

Nonparametric Bayesian analysis for support boundary recovery

Markus Reiß and Johannes Schmidt-Hieber

Abstract

Given a sample of a Poisson point process with intensity $\lambda_f(x, y) = n\mathbf{1}(f(x) \leq y)$, we study recovery of the boundary function f from a nonparametric Bayes perspective. Because of the irregularity of this model, the analysis is non-standard. We derive contraction rates with respect to the L^1 -norm for several classes of priors, including Gaussian priors, priors based on (truncated) random series, compound Poisson processes, and subordinators. We also investigate the limiting shape of the posterior distribution and derive a nonparametric version of the Bernstein-von Mises theorem for a specific class of priors on a function space with increasing parameter dimension. We show that the marginal posterior of the functional $\vartheta = \int f$ does some automatic bias correction and contracts with a faster rate than the MLE. In this case, $1 - \alpha$ -credible sets are also asymptotic $1 - \alpha$ confidence intervals. It is also shown that the frequentist coverage of credible sets is lost under model misspecification.

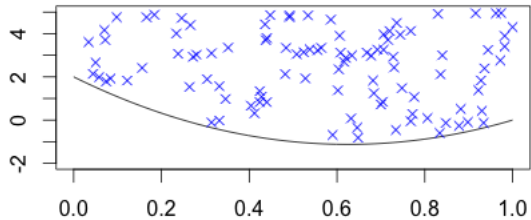
1 Introduction

We consider a support boundary detection model, where a Poisson point process (PPP) N on $[0, 1] \times \mathbb{R}$ is observed with intensity

$$\lambda(x, y) = \lambda_f(x, y) = n\mathbf{1}(f(x) \leq y).$$

The statistical task is to recover the unobserved lower boundary $f : [0, 1] \rightarrow \mathbb{R}$ of the support of λ . A simulated dataset is displayed below. This boundary detection model can be seen as a continuous analog of the nonparametric regression model with discrete equidistant design and exponential errors, that is, we observe $Y_{i,n} = f(i/n) + \varepsilon_{i,n}$, $i = 1, \dots, n$, and $(\varepsilon_{i,n})_i$ are i.i.d. exponential random variables, cf. [24, 15]. Due to the one-sided error distribution, this model, with f in a parametric class, is not Hellinger differentiable and therefore irregular.

Frequentist estimation of the object boundary in an image is a well-studied problem and the optimal minimax estimation rates over standard function classes are known. It has also been discovered that for estimation of functionals the maximum likelihood estimator (MLE) is not rate-optimal and some correction term has to be introduced. Since this is the main motivation for our



Simulated dataset (blue) of Poisson point process model with boundary function $f(x) = 8(x - 1/2)^2 - 2x$ (black) and $n = 20$.

Bayesian analysis, let us describe this phenomenon in more detail. For some parameter spaces Θ , such as Hölder balls with smoothness index $\beta \leq 1$, the nonparametric MLE exists. For pointwise estimation, the MLE achieves moreover the minimax estimation rate. For Hölder balls of β -smooth functions this rate is $n^{-\beta/(\beta+1)}$, cf. [19]. If we, however, estimate a functional such as $\vartheta = \int f$, then the MLE $\hat{\vartheta}^{\text{MLE}} = \int \hat{f}^{\text{MLE}}$ converges with rate $n^{-\beta/(\beta+1)}$ which is slower than the minimax estimation rate $n^{-(\beta+1/2)/(\beta+1)}$. A rate-optimal estimator can be obtained if we subtract a term that scales with the number of observations lying on the boundary of the MLE and consider

$$\hat{\vartheta} = \int \hat{f}^{\text{MLE}} - \frac{\text{number of data points } (X_i, Y_i) \text{ on the boundary of } \hat{f}^{\text{MLE}}}{n},$$

cf. [28]. This can be viewed as a bias corrected version of the MLE that accounts for the fact that \hat{f}^{MLE} overshoots the true boundary function f .

It is then natural to ask whether a Bayesian approach would do this correction automatically, by concentrating the posterior mass around $\hat{\vartheta}$. In this work we show that the answer to this problem depends on the choice of the prior. As a positive result, we prove that in a simple nonparametric model and for a prior that is natural for the corresponding parameter space, the posterior concentrates around $\hat{\vartheta}$ with the optimal contraction rate. In this setting, the posterior corrects the MLE by the right amount. Moreover, by deriving a Bernstein-von Mises theorem, we can show that $1 - \alpha$ -credible sets are also asymptotic $1 - \alpha$ -confidence sets for the functional ϑ . As a negative example, we consider a setting where the posterior contracts to the true function f with the optimal rate, but the bias correction of the marginal posterior for $\vartheta = \int f$ is of the wrong order. In this case, we can further show that credible sets have asymptotically no frequentist coverage. As a conclusion, these results indicate that only for very specific priors the Bayesian method can undo the bias induced by the MLE.

Besides an improved understanding of the uncertainty quantification of the Bayes method,

the goal of this article is to describe more generally the frequentist properties of the posterior for the support boundary detection model. In a first part, we therefore study posterior contraction for the boundary function f . As loss function, we consider the L^1 -distance which is linked to the information geometry of the support boundary detection model. Posterior contraction for the Hellinger loss is well-studied and can be reduced to conditions on the entropy of the parameter space and the small ball probability of the prior, cf. [9, 10]. We derive a modification of this result which is applicable for the support boundary detection model. Related to that, we show the following surprising result. If the posterior is restricted to functions that lie below the true function, then posterior contraction follows already from the behavior of the one-sided small ball prior probability. In this case no bound on the entropy is necessary. On the contrary, for functions which lie above the true function, we essentially only need the entropy bound.

Given the general contraction result, we apply this to several classes of priors. In a first step, we study Gaussian priors and derive an analog of the result in [34] for the support boundary detection model. We also study posterior contraction for random wavelet series priors with independent but not necessarily Gaussian random coefficients. For these priors we derive a result on small ball probabilities which is of independent interest. The corresponding contraction rates only match with the minimax estimation rates for one smoothness index. Below this specific smoothness the contraction rates can be improved if more heavy-tailed distributions on the wavelet coefficients are used. We also prove that truncated random wavelet series priors achieve the adaptive rates up to logarithmic factors.

A considerable amount of work is spent on understanding posterior contraction with compound Poisson process (CPP) and subordinator priors. These are interesting classes of priors which received only little attention in nonparametric Bayes theory so far. Neglecting logarithmic factors, we prove that the CPP prior with fixed intensity can adapt to the smoothness of the signal up to smoothness one and achieves the minimax estimation rate for piecewise constant functions with increasing number of pieces. We also consider subordinator priors which are natural for spaces of increasing functions. It is proved that under general assumptions the contraction rate is $(\log n/n)^{1/2}$ and we show that for specific priors the $\log n$ term can be avoided.

In the last part, we study Bernstein-von Mises theorems for the function f and the marginal posterior on $\vartheta = \int f$. For that we consider a prior on the piecewise constant functions with an increasing number of pieces. Depending on the smoothness of the prior, we determine a maximal growth of the number of pieces such that the prior influence vanishes in the limit and derive an explicit expression for the approximating distribution. Due to the irregularity of the model, the limiting shape of the posterior is non-Gaussian and also not centered at the

MLE. This is thus a non-standard Bernstein-von Mises theorem. The marginal posterior on $\vartheta = \int f$ follows asymptotically a normal distribution centered at the bias corrected MLE. There are moreover intervals which are simultaneously asymptotic $1 - \alpha$ credible and confidence intervals with length shrinking at the optimal rate. The Bayesian approach clearly outperforms the MLE in this case. Under model misspecification, however, credible sets can have asymptotically no frequentist coverage.

Bayesian methods for irregular or boundary detection problems have attracted considerable attention especially because the MLE approach is often inefficient. [7] compares Bayes estimators with the MLE in a parametric model that is irregular. In [2] a Bernstein-von Mises theorem is derived for parameters which are on the boundary of the parameter space. The limit distribution consists in this case of Gaussian and exponentially distributed components. [17] considers posterior contraction around θ given i.i.d. observations from a class of nonparametric densities of the form $\eta(x - \theta)$ with $\eta(y) = 0$ for $y < 0$ and $\eta(y) > 0$ for $y \geq 0$. This can be viewed as a semiparametric, irregular model, where the nuisance parameter is the unknown distribution of the noise. This article also contains a Bernstein-von Mises result in a parametric irregular model. For nonparametric models, [23] considers Bayesian methods for Poisson point processes, but does not cover boundary detection. While writing the article, we became aware of the recent work by Li and Ghoshal [22]. They consider a nonparametric Bayes approach for detection of a closed boundary of an object in an image assuming different distributions of the response variable inside and outside the object in the image. Compared with the model considered in this article, their model is regular and the likelihood ratios are always well-defined. Although the models are not the same, there are various similarities such as the underlying L^1 -geometry. Both articles derive contraction rates (although for different classes of priors). Otherwise [22] focuses more on computational issues while our main motivation are non-standard Bernstein-von Mises theorems and frequentist coverage of credible sets.

The paper is structured as follows. In Section 2, we derive a general result relating posterior contraction to entropy and small ball estimates. This result is then used in Section 3 to derive a criterion for posterior contraction under Gaussian priors. Section 4 studies wavelet expansion priors. Compound Poisson process priors and subordinator priors are investigated in Section 5. Bernstein -von Mises type theorems and results on the frequentist coverage of credible sets can be found in Section 6. Technicalities and proofs are deferred to an appendix.

Notation: We write $(x)_+ = \max(x, 0)$ and denote the indicator function of a set A by $\mathbf{1}_A = \mathbf{1}(\cdot \in A)$. For $p \in [1, \infty]$, $\|\cdot\|_p$ denotes the $L^p[0, 1]$ -norm. Inequalities in L^1 -norm are assumed to hold almost everywhere. Let $\lfloor \beta \rfloor$ denote the largest integer strictly smaller than $\beta > 0$.

The β -Hölder norm is $\|f\|_{C^\beta} := \sum_{j=1}^{\lfloor \beta \rfloor} \|f^{(j)}\|_\infty + \sup_{x \neq y} |f^{(\lfloor \beta \rfloor)}(x) - f^{(\lfloor \beta \rfloor)}(y)|/|x - y|^{\beta - \lfloor \beta \rfloor}$. We denote by $C^\beta(R)$ the class of functions f on $[0, 1]$ with $\|f\|_{C^\beta} \leq R$. We further write $N = \sum_i \delta_{(X_i, Y_i)}$ for a random measure on $[0, 1] \times \mathbb{R}$ and often identify N with its support points $(X_i, Y_i)_i$. For two positive sequences $(a_n)_n, (b_n)_n$ we write $a_n \lesssim b_n$ if there is a constant C such that $a_n \leq Cb_n$ for all n . If $a_n \lesssim b_n$ and $b_n \lesssim a_n$ then we write $a_n \asymp b_n$.

2 General results on posterior contraction rates

2.1 Likelihood and Bayes formula

Before stating the main result on posterior contraction, we first study the likelihood in the support boundary detection model. From that we can then derive expressions for the information distances and a specific form of the Bayes formula.

Denote by $P_f = P_f^n$ the distribution of a PPP with intensity measure $\Lambda_f(B) = \int_B \lambda_f$ for Borel sets B in $[0, 1] \times \mathbb{R}$ with Lebesgue density $\lambda_f(x, y) = n \mathbf{1}(f(x) \leq y)$, where f is some function in $L^1([0, 1])$. The likelihood ratio dP_f/dP_g is only defined for $g \leq f$, as otherwise P_g does not dominate P_f . The fact that the observation laws are not necessarily mutually absolutely continuous is a distinctive feature of support estimation problems and will play a major role in the analysis. Recall that for a Poisson point process $(X_i, Y_i)_i$ denote the coordinates of the support points in $[0, 1] \times \mathbb{R}$.

2.1 Lemma. *For $g \leq f$ the likelihood ratio has the explicit form*

$$\frac{dP_f}{dP_g} = \exp \left(n \int_0^1 (f - g)(x) dx \right) \cdot \mathbf{1}(\forall i : f(X_i) \leq Y_i). \quad (2.1)$$

The information geometry of the model is driven by the $L^1([0, 1])$ -norm $\|\cdot\|_1$. Indeed, the Hellinger affinity is $\rho(P_f, P_g) = \int \sqrt{dP_f dP_g} = \exp(-\frac{n}{2} \|f - g\|_1)$. This implies for the squared Hellinger distance

$$H^2(P_f, P_g) = 2 - 2\rho(P_f, P_g) = 2 - 2 \exp \left(-\frac{n}{2} \|f - g\|_1 \right) \leq n \|f - g\|_1, \quad \forall f, g \in L^1([0, 1]).$$

Similarly, the Kullback-Leibler divergence satisfies $K(P_f, P_g) = n \|f - g\|_1$ if $g \leq f$ and $K(P_f, P_g) = \infty$ otherwise.

As priors, we consider the distributions Π of stochastic processes $(X_t)_{t \in [0, 1]}$ on a Polish space (Θ, d) equipped with its Borel σ -algebra. We aim for a Bayes formula of the form

$$\Pi(B|N) = \frac{\int_B \frac{dP_f}{dP_{f_0}}(N) d\Pi(f)}{\int_\Theta \frac{dP_f}{dP_{f_0}}(N) d\Pi(f)}. \quad (2.2)$$

Since in the boundary detection model the likelihood ratio does in general not exist, the formula has to be modified. The next result provides us with a Bayes formula under the frequentist assumption that the data are generated from P_{f_0} .

2.2 Lemma. *For $f_0 \in L^1([0, 1])$, a prior Π on the Polish space Θ and a Borel set $B \subset \Theta$ we have an explicit Bayes formula under the law P_{f_0} :*

$$\Pi(B|N) = \frac{\int_B e^{n \int f} \mathbf{1}(\forall i : f(X_i) \leq Y_i) d\Pi(f)}{\int_\Theta e^{n \int f} \mathbf{1}(\forall i : f(X_i) \leq Y_i) d\Pi(f)} = \frac{\int_B e^{-n \int (f_0 - f)_+} \frac{dP_{f \vee f_0}}{dP_{f_0}}(N) d\Pi(f)}{\int_\Theta e^{-n \int (f_0 - f)_+} \frac{dP_{f \vee f_0}}{dP_{f_0}}(N) d\Pi(f)} \quad P_{f_0}\text{-a.s.}$$

The right-hand side is well-defined since $dP_{f \vee f_0}/dP_{f_0}$ exists. Compared to (2.2), the likelihood ratios are reweighted in the Bayes formula by a factor $e^{-n \int (f_0 - f)_+}$. In particular, if $f \leq f_0$, the integrands are equal to the deterministic values $e^{-n \int (f_0 - f)}$.

2.2 Main results

We state the main theorem reducing posterior contraction to conditions on the entropy and small ball probabilities. The result is an analog of the general contraction theorems in [9, 10]. Denote by $N(\varepsilon, \mathcal{F}, d)$ the ε -covering number of \mathcal{F} with respect to the distance d .

2.3 Theorem. *If for some $\Theta_n \subset \Theta$, some rate $\varepsilon_n \rightarrow 0$ and constants $C, C', C'' \geq 1$, $A > 0$*

- (i) $N(\varepsilon_n, \Theta_n, \|\cdot\|_\infty) \leq C'' e^{C' n \varepsilon_n};$
- (ii) $\Pi(f : \|f - f_0\|_1 \leq A \varepsilon_n, f \leq f_0) \geq e^{-C n \varepsilon_n};$
- (iii) $\Pi(\Theta_n^c) \leq C'' e^{-(C+A+1)n \varepsilon_n},$

then, there exists a constant M such that

$$E_{f_0} [\Pi(f : \|f - f_0\|_1 \geq M \varepsilon_n | N)] \leq 3C'' e^{-n \varepsilon_n}.$$

Condition (i) can be relaxed to any of the conditions of Corollary 2.6.

For (ii) we need a lower bound on the one-sided small ball probabilities. A stronger version of (ii) which is often easier to verify is given by

$$(ii)': \quad \Pi(f : \|f + A \varepsilon_n / 2 - f_0\|_\infty \leq A \varepsilon_n / 2) \geq e^{-C n \varepsilon_n}. \quad (2.3)$$

The proof of the theorem is deferred to the appendix, yet main intermediate results are presented here. It will be convenient to establish posterior contraction for $\int (f_0 - f)_+$ and $\int (f - f_0)_+$ separately. Surprisingly, for posterior contraction with respect to $\int (f_0 - f)_+$ we

only need the small ball estimate of the prior probability, but no bound on the entropy. In contrast, posterior contraction for $\int(f - f_0)_+$ only requires that (i) and (iii) of Theorem 2.3 hold.

2.4 Proposition. *If for some constants $C > 0, A \geq 1$*

$$\Pi\left(f : \int(f_0 - f) \leq A\varepsilon_n, f \leq f_0\right) \geq e^{-Cn\varepsilon_n},$$

then

$$E_{f_0}\left[\Pi\left(f : \int(f_0 - f)_+ \geq (1 + A + C)\varepsilon_n \middle| N\right)\right] \leq e^{-n\varepsilon_n}.$$

The one-sided small ball probability can be viewed as a prior mass condition on a Kullback-Leibler ball in view of $\{f : \text{KL}(P_{f_0}, P_f) \leq A\varepsilon_n n\} = \{f : \int(f_0 - f) \leq A\varepsilon_n, f \leq f_0\}$. To establish posterior contraction with respect to the loss $\int(f - f_0)_+$, we need to understand the testing theory in the boundary detection model, which is non-standard due to the lack of absolute continuity in general. The likelihood ratio test $\phi = \mathbf{1}(dP_g/dP_{f \wedge g} \geq dP_f/dP_{f \wedge g})$ behaves well for testing f against g :

$$E_f\phi + E_g[1 - \phi] = \int \left(\frac{dP_f}{dP_{f \wedge g}} \wedge \frac{dP_g}{dP_{f \wedge g}} \right) dP_{f \wedge g} \leq \rho(P_f, P_g) = e^{-\frac{n}{2}\|f - g\|_1}.$$

Robustness with respect to the Hellinger distance, however, in the sense that for some $\alpha, \beta > 0$, and all n ,

$$E_f\phi + \sup_{h: \|h - g\|_1 \leq \alpha\|f - g\|_1} E_h[1 - \phi] \leq e^{-\beta n\|f - g\|_1}$$

holds, is violated: if $f \leq g \wedge h$, by direct computations, $E_h[1 - \phi] = P_h(\exists i : g(X_i) > Y_i) = 1 - e^{-n\int(g - h)_+}$, which is generally much larger than $e^{-\beta n\|f - g\|_1}$. Yet, under the additional assumption $g \leq h$ the type II error vanishes completely and we find for $f \leq g$

$$E_f\phi \leq e^{-\frac{n}{2}\|f - g\|_1} \quad \text{and} \quad \sup_{h \geq g} E_h[1 - \phi] = 0.$$

To control the posterior, it is therefore natural to use one-sided bracketing entropy. Consider a subset \mathcal{F} of $L^1([0, 1])$. The one-sided bracketing number $N_{[\cdot]}(\delta, \mathcal{F})$ is the smallest number M of functions $\ell_1, \dots, \ell_M \in L^1([0, 1])$ such that for any $f \in \mathcal{F}$ there exists $j \in \{1, \dots, M\}$ with $\ell_j \leq f$ and $\int(f - \ell_j) \leq \delta$. For some function f_0 and integer n consider the separation quantity

$$S_{[\cdot]}(n, \mathcal{F}, f_0) = \inf_{(\ell_j)_{j \in J}} \sum_{j \in J} e^{-n\int(\ell_j - f_0)_+} \in [0, \infty],$$

where the infimum is taken over (not necessarily finite) subsets J of the integers and functions $(\ell_j)_{j \in J} \subset L^1([0, 1])$ such that for any $f \in \mathcal{F}$ there exists $j \in J$ with $\ell_j \leq f$. In both definitions the functions ℓ_j are not required to be in \mathcal{F} .

In view of the next result, the quantity S_{\lceil} , which can be seen as a weighted covering number, is the natural complexity measure for Θ .

2.5 Proposition. *If $\Pi(f : f \leq f_0) > 0$, then for any Borel set $B \subseteq \Theta$*

$$E_{f_0}[\Pi(f \in B | N)] \leq S_{\lceil}(n, B, f_0).$$

Notice that the right-hand side does not depend on the prior. Weighted covering numbers might be small even for non-compact parameter spaces (see the proof of Proposition 5.6 for an example) and have been used before in nonparametric Bayes theory, cf. [14], Section 4. For many specific problems, covering or bracketing numbers are sufficient and we can further upper bound the right-hand side in Proposition 2.5, using $-\int(\ell_j - f_0)_+ \leq -\int(f - f_0)_+ + \int(f - \ell_j)$ which implies that for any $\Theta_n \subset \Theta$,

$$S_{\lceil}(n, \{f \in \Theta_n : \int(f - f_0)_+ \geq \varepsilon\}, f_0) \leq e^{-n\varepsilon/2} N_{\lceil}(\varepsilon/2, \Theta_n) \leq e^{-n\varepsilon/2} N(\varepsilon/4, \Theta_n, \|\cdot\|_{\infty}).$$

2.6 Corollary. *Work under the assumptions of Proposition 2.5. If $C \geq 1$, then*

$$E_{f_0}\left[\Pi\left(f \in \Theta_n : \int(f - f_0)_+ \geq 4C\varepsilon_n \mid N\right)\right] \leq C'' e^{-n\varepsilon_n}$$

holds under any of the following conditions:

- (i) $S_{\lceil}(n, \{f \in \Theta_n : \int(f - f_0)_+ \geq 4C\varepsilon_n\}, f_0) \leq C'' e^{-n\varepsilon_n};$
- (ii) $N_{\lceil}(\varepsilon_n, \Theta_n) \leq C'' e^{Cn\varepsilon_n};$
- (iii) $N(C\varepsilon_n, \Theta_n, \|\cdot\|_{\infty}) \leq C'' e^{Cn\varepsilon_n}.$

We can avoid the entropy condition if we control instead the risk of an estimator. Indeed, $\inf_{\phi} E_{\theta_0} \phi + \sup_{\theta \in \Theta : \ell(\theta, \theta_0) \geq 2\varepsilon} E_{\theta} [1 - \phi] \leq 2 \inf_{\hat{\theta}} \sup_{\theta \in \Theta} P_{\theta}(\ell(\hat{\theta}, \theta) \geq \varepsilon)$ which follows by studying the test $\phi = \mathbf{1}(\ell(\hat{\theta}, \theta) \geq \varepsilon)$ given an estimator $\hat{\theta}$. If the nonparametric MLE for f exists, we have a particularly simple relation in the support boundary detection model between posterior contraction of $\int(f - f_0)_+$ and the excess probability of the MLE.

2.7 Lemma. *Assume that $\Theta_n \subseteq \Theta$ contains f_0 and is closed under taking maxima. If the maximum likelihood estimator \hat{f}^{MLE} , based on the parameter space Θ_n , exists, then*

$$E_{f_0}\left[\Pi\left(f \in \Theta_n : \int(f - f_0)_+ > \varepsilon_n \mid N\right)\right] \leq P_{f_0}\left(\int(\hat{f}^{\text{MLE}} - f_0)_+ > \varepsilon_n\right). \quad (2.4)$$

As in the proof of Proposition 2.5 the upper bound is independent of the prior. It is well-known that posterior contraction with rate ε_n implies existence of a frequentist estimator with rate of convergence ε_n , cf. Theorem 2.5 in [9]. Inequality (2.4) shows that also the other

direction may hold, namely that convergence of an estimator implies posterior contraction with the same rate. Regarding the assumptions, a sufficient condition for the existence of the MLE is that Θ is closed under arbitrary maxima: $f_i \in \Theta, i \in I \Rightarrow \bigvee_{i \in I} f_i \in \Theta$. Examples of function spaces which are closed under the maximum are Hölder balls, monotone functions and convex functions.

3 Gaussian process priors

A common choice in nonparametric Bayes is to pick the distribution of a Gaussian process as prior probability measure. Given a Gaussian process prior Π , the seminal work in [34] relates posterior contraction to the small ball prior probability and approximation properties in the reproducing kernel Hilbert space (RKHS) generated by Π . The following result adapts Theorem 2.1 in [34] to our setting.

3.1 Theorem. *Consider as prior the distribution of a Gaussian process X with sample paths in the space $(\mathcal{C}[0, 1], \|\cdot\|_\infty)$. Write $\|\cdot\|_{\mathbb{H}}$ for the RKHS-norm induced by the covariance operator of X . If $\varepsilon_n \geq n^{-1}$ and*

$$\inf_{h: \|h+2\varepsilon_n-f_0\|_\infty \leq \varepsilon_n} \|h\|_{\mathbb{H}}^2 - \log \mathbb{P}(\|X\|_\infty \leq \varepsilon_n) \leq n\varepsilon_n, \quad (3.1)$$

then there exists a constant M such that for all n

$$E_{f_0}[\Pi(f : \|f - f_0\|_1 \geq M\varepsilon_n | N)] \leq 3e^{-n\varepsilon_n}.$$

Condition (3.1) is slightly different compared to (1.2) and (1.3) in [34]. As a bound we have $n\varepsilon_n$ instead of $n\varepsilon_n^2$ and in the RKHS part there is an extra term $2\varepsilon_n$ which accounts for the one-sided prior mass condition in Theorem 2.3.

As an example let us study the Brownian motion prior with a random starting value. The prior is the law of the process $(X_0 + W_t)_{t \in [0,1]}$ with W_t a Brownian motion and X_0 an independent standard normal random variable. Arguing as in [34], Section 4.1, we find for the corresponding RKHS norm $\|h\|_{\mathbb{H}}^2 = \|h'\|_{L^2[0,1]}^2 + h(0)^2$, $\Pi(f : \|f\|_\infty \leq \varepsilon_n) \geq \mathbb{P}(|X_0| \leq \varepsilon_n/2) \mathbb{P}(\|W\|_\infty \leq \varepsilon_n/2) \gtrsim \varepsilon_n e^{-C/\varepsilon_n^2}$ and $\inf_{h: \|h+2\varepsilon_n-f_0\|_\infty \leq \varepsilon_n} \|h\|_{\mathbb{H}}^2 \lesssim \varepsilon_n^{2-2/\beta}$. The L^1 -contraction rate is therefore

$$\begin{cases} n^{-\frac{\beta}{2-\beta}}, & \text{for } \beta \leq 1/2, \\ n^{-\frac{1}{3}}, & \text{for } \beta \geq 1/2. \end{cases}$$

This coincides with the minimax rate $n^{-\beta/(\beta+1)}$ if $\beta = 1/2$. For $\beta > 1/2$, we do not gain anything more in the contraction rate by imposing more smoothness on the signal.

For $\beta < 1/2$ the rate is slower than the minimax rate. This behavior of the posterior for Gaussian priors is well-known in nonparametric Bayes theory.

4 Wavelet expansion priors

Series expansions provide a natural way to construct priors on function spaces. In this section, we study process priors $(X_t)_{t \in [0,1]}$ which admit a wavelet expansion

$$X_t = \sum_{j,k} d_{j,k} \xi_{j,k} \psi_{j,k}(t) \quad \text{in } L^2[0,1]. \quad (4.1)$$

Here, $d_{j,k}$ are real numbers and $\xi_{j,k}$ are i.i.d. random variables with Lebesgue density f_ξ . As a prior on the function f this means that each wavelet coefficient of f is drawn independently from the distribution of $d_{j,k} \xi_{j,k}$. For convenience, we restrict ourselves in this section to s -regular, boundary corrected and compactly supported wavelet bases $(\psi_{j,k})$ in $L^2([0,1])$ as constructed in Section 4 of [8].

Wavelet expansion priors have been studied in different nonparametric models with uniform random variables $\xi_{j,k}$, cf. [11, 26]. Moreover, [35] derives bounds on the small ball probabilities of Gaussian processes of the form (4.1). Below, we derive posterior contraction rates for a class of distributions $\xi_{j,k}$. To start with, we prove the following general lower bound on small ball probabilities, which is of independent interest.

4.1 Lemma. *Assume (4.1) with symmetric and unimodal f_ξ and $|d_{j,k}| \asymp 2^{-\frac{j}{2}(2\alpha+1)}$ for some $\alpha > 0$. Suppose further that there are constants L and $\delta > 0$ such that*

$$\mathbb{E}[|\xi_{j,k}|^{(1+\delta)/\alpha}] \leq L.$$

Then for all $\beta \in (0, s]$, $R > 0$ there exists a constant D such that

$$\inf_{h \in \mathcal{C}^\beta(R)} \mathbb{P}(\|X - h\|_\infty \leq \varepsilon) \geq f_\xi(D\varepsilon^{-(\alpha-\beta)_+/\beta})^{D\varepsilon^{-1/(\alpha \wedge \beta)}} \quad \text{for all } 0 < \varepsilon \leq 1.$$

For $\beta \geq \alpha$ the lower bound is of the order $e^{-D\varepsilon^{-1/\alpha}}$. For $\beta < \alpha$ the lower bound depends on the tails of the distribution: heavier tails lead to larger lower bounds on the small ball probabilities and in consequence to better contraction rates. The result implies that the contraction rate in Theorem 2.3 is at best $\varepsilon_n = n^{-\alpha/(1+\alpha)}$, which is the solution of the equation $\varepsilon_n^{-1/\alpha} = n\varepsilon_n$.

4.2 Theorem. *Consider the process in (4.1) as prior with symmetric and unimodal f_ξ and $|d_{j,k}| \asymp 2^{-\frac{j}{2}(2\alpha+1)}$ for some $\alpha > 1$. Suppose that for some $\gamma > 0$, $f_\xi(x) \leq \gamma^{-1}e^{-\gamma|x|}$ for all*

$x \in \mathbb{R}$ and fix $\beta \in (0, s]$, $R > 0$. For any sequence $\varepsilon_n \rightarrow 0$, satisfying

$$n\varepsilon_n^{\frac{1+(\alpha\wedge\beta)}{\alpha\wedge\beta}} \asymp -\log f_\xi\left(D\varepsilon_n^{-\frac{(\alpha-\beta)_+}{\beta}}\right),$$

there exist constants M and c such that for all $f_0 \in \mathcal{C}^\beta(R)$ and all n

$$E_{f_0}[\Pi(f : \|f - f_0\|_1 \geq M\varepsilon_n | N)] \leq e^{-cn\varepsilon_n}.$$

The assumption $\alpha > 1$ is imposed in order to bound the bracketing entropy in Theorem 2.3. One of the consequences of Theorem 4.2 is that the posterior contracts faster in the regime $\beta < \alpha$ if heavier-tailed distributions are used. This is illustrated by two specific examples.

4.3 Example. (A) If $\xi_{j,k} \sim \mathcal{N}(0, 1)$, we find for a sufficiently large constant C

$$\inf_{h \in \mathcal{C}^\beta(R)} \mathbb{P}(\|X - h\|_\infty \leq \varepsilon) \geq \exp\left(-C\varepsilon^{-\left(\frac{1+2\alpha-2\beta}{\beta} \vee \frac{1}{\alpha}\right)}\right).$$

For $\alpha = 1/2$, the bound becomes $\exp(-C\varepsilon^{-2(\frac{1-\beta}{\beta} \vee 1)})$, which is the same as for the Brownian motion prior. The resulting posterior contraction rate is

$$\varepsilon_n = n^{-\frac{\beta \wedge \alpha}{1+\alpha+(\alpha-\beta)_+}}.$$

(B) If $\xi_{j,k}$ follows a Laplace (double-exponential) distribution, we obtain

$$\inf_{h \in \mathcal{C}^\beta(R)} \mathbb{P}(\|X - h\|_\infty \leq \varepsilon) \geq \exp\left(-C\varepsilon^{-\left(\frac{1+\alpha-\beta}{\beta} \vee \frac{1}{\alpha}\right)}\right).$$

The posterior contraction rate becomes $\varepsilon_n = n^{-\frac{\beta \wedge \alpha}{1+\alpha}}$ which improves the rate in (A) for $\beta < \alpha$.

We can also obtain a fully adaptive result (up to $\log n$ factors) using a random truncation of the wavelet expansion prior. The prior can be realized via a hierarchical construction. In a first step, we draw the maximal resolution level J from a distribution satisfying $P(J = j) \propto 2^{-j}$. Given J , generate

$$X_t = \sum_{j \leq J, k} \xi_{j,k} \psi_{j,k}(t) \quad (4.2)$$

with $\psi_{j,k}$ as in (4.1) and $(\xi_{j,k})_{j,k}$ an i.i.d. sequence of random variables with positive and continuous Lebesgue density f_ξ . In this prior the regularization is induced by the truncation of the wavelet series and compared with (4.1) we can set $d_{j,k} = 1$.

4.4 Lemma. Consider the random truncation prior (4.2). For $\beta \in (0, s]$, $R > 0$ there exists a constant D such that

$$\inf_{h \in \mathcal{C}^\beta(R)} \mathbb{P}(\|X - h\|_\infty \leq \varepsilon) \geq \varepsilon^{D\varepsilon^{-1/\beta}} \quad \text{for all } 0 < \varepsilon \leq 1.$$

4.5 Theorem. Consider the random truncation prior (4.2). Suppose that for some $\gamma > 0$, $f_\xi(x) \leq \gamma^{-1}e^{-\gamma|x|}$ for all $x \in \mathbb{R}$ and fix $\beta \in (0, s]$, $R > 0$. Then there exist constants M and c such that for all $f_0 \in \mathcal{C}^\beta(R)$ and n

$$E_{f_0}[\Pi(f : \|f - f_0\|_1 \geq M\varepsilon_n | N)] \leq e^{-cn\varepsilon_n}$$

with

$$\varepsilon_n = \left(\frac{\log n}{n} \right)^{\frac{\beta}{\beta+1}}.$$

A key ingredient of the proof is a connection to Besov spaces which allows us to obtain entropy bounds.

5 Compound Poisson and subordinator priors

5.1 Compound Poisson process (CPP) priors

In this section we study posterior contraction for compound Poisson process priors defined on the space $\Theta = D[0, 1]$ of càdlàg functions, equipped with the Skorokhod topology. A compound Poisson process Y on $[0, 1]$ can be written as $Y_t = \sum_{i=1}^{N_t} \Delta_i$ with a Poisson process $(N_t)_{t \geq 0}$ of intensity $\lambda > 0$ and an i.i.d. sequence (Δ_i) of random variables, independent of the Poisson process. We denote the distribution of Δ_1 by G . We randomize the starting value and consider

$$X_t = \Delta_0 + \sum_{i=1}^{N_t} \Delta_i = \sum_{i=0}^{N_t} \Delta_i, \quad (5.1)$$

where $\Delta_0 \sim G$, independently of $(\Delta_i)_{i \geq 1}$ and $(N_t)_{t \geq 0}$.

A natural parameter space for CPP priors are piecewise constant functions with $K = K_n$ jump points

$$\text{PC}(K, R) = \left\{ f : f = \sum_{j=1}^K a_j \mathbf{1}_{[t_{j-1}, t_j)} \text{ with } 0 := t_0 < t_1 < \dots < t_K := 1, |a_j| \leq R \right\}.$$

We are interested in the case where the number of jump points K_n grows with n . In this case we have $2K_n - 1$ parameters. Since the squared parametric rate is n^{-1} , we expect the best possible contraction rate to be K_n/n .

A CPP can also be viewed as a hierarchical prior on f in the spirit of [6, 5]. The hierarchical CPP construction picks in a first step a model dimension prior $\pi \sim \text{Pois}(\lambda)$. Given model dimension k , the jumps points $\mathbf{t} = (t_1, \dots, t_k)$ are chosen uniformly at random from $[0, 1]^k$. The last step is then to assign jump sizes $\mathbf{b} = (b_0, b_1, \dots, b_k)$ with b_j drawn independently from the jump distribution G . This generates random càdlàg functions $f = \sum_{j=0}^k b_j \mathbf{1}_{[t_j, 1]}$ with $t_0 := 0$. We can write the CPP prior in closed form as a prior on k, \mathbf{t} , and \mathbf{b}

$$(k, \mathbf{t}, \mathbf{b}) \mapsto e^{-\lambda} \frac{\lambda^k}{k!} \prod_{j=0}^k g(b_j) \mathbf{1}(\mathbf{t} \in [0, 1]^k). \quad (5.2)$$

Since λ is fixed, for most draws of the prior the number of jumps will be of order λ . As we show below, the CPP prior puts still enough mass around functions with an increasing number of jumps to ensure nearly optimal posterior contraction rates for Hölder functions and for piecewise constant function with an increasing number of pieces. Let us also mention that the CPP prior randomizes over the jump points and should therefore be able to adapt to local smoothness. This might be an advantage compared with random histogram priors where the function jumps on a fixed uniform grid.

5.1 Lemma. *Let X be the randomly initialized CPP in (5.1). Suppose that the jump distribution G has a positive and continuous Lebesgue density g on \mathbb{R} . Then:*

(i) *for $0 < \beta \leq 1$ there exists a positive constant $c = c(\beta, R)$ such that*

$$\inf_{f \in \mathcal{C}^\beta(R)} \mathbb{P}(\|X - f\|_\infty \leq \varepsilon) \geq e^{-2\lambda} (1 \wedge \lambda)^{(2R/\varepsilon)^{1/\beta}} \varepsilon^{c\varepsilon^{-1/\beta}} \quad \text{for all } 0 < \varepsilon \leq \frac{R\lambda}{2};$$

(ii) *there exists a positive constant $c = c(R)$ such that*

$$\inf_{f \in \text{PC}(K_n, R)} \mathbb{P}(\|X - f\|_1 \leq \varepsilon, X \leq f) \geq e^{-\lambda} (1 \wedge \lambda)^{K_n} \left(\frac{\varepsilon}{K_n} \right)^{cK_n}, \quad \text{for all } 0 < \varepsilon \leq \frac{1}{2}.$$

5.2 Lemma. *Consider the randomly initialized CPP (5.1) and assume that there are constants $\gamma, L > 0$ such that $\mathbb{P}(|\Delta_i| \geq s) \leq L^{-1} e^{-Ls^\gamma}$ for all $s \geq 0$. Then for any $M, \varepsilon > 0$ and any sequence $(K_n)_n$ of positive integers there exist Borel sets $(\Theta_n)_n$ and constants C', C'' , such that for all n*

$$\mathbb{P}(X \notin \Theta_n) \leq C'' K_n^{-MK_n} \quad \text{and} \quad N_{[\varepsilon, \Theta_n, \|\cdot\|_1]} \leq C'' \left(\frac{K_n}{\varepsilon} \right)^{C'K_n}.$$

To derive contraction rates, we may now apply Theorem 2.3 with (i) replaced by Condition (ii) in Corollary 2.6. The previous lemma allows us to pick Θ_n such that the entropy is of the right order and (iii) of Theorem 2.3 is satisfied. Indeed $N_{[]}(\varepsilon_n, \Theta_n, \|\cdot\|_1) \leq C''' e^{Cn\varepsilon_n}$ and $\Pi(\Theta_n^c) \leq e^{-cMn\varepsilon_n}$ if

$$K_n \log \left(\frac{K_n}{\varepsilon_n} \right) \leq n\varepsilon_n \quad \text{and} \quad K_n \log K_n \geq cn\varepsilon_n.$$

If $\varepsilon_n \gtrsim n^{\rho-1}$ for some $\rho > 0$ and n is sufficiently large, then $\log(n\varepsilon_n)/\log n$ remains positive and $K_n = n\varepsilon_n/\log n$ satisfies both inequalities for some $c > 0$.

5.3 Theorem. *Consider the prior in (5.1). Under the joint assumptions of Lemma 5.1 and Lemma 5.2 there exist constants M and c such that*

$$\sup_{f_0 \in \mathcal{C}^\beta(R)} E_{f_0} \left[\Pi \left(f : \|f - f_0\|_1 \geq M(\log n/n)^{\beta/(1+\beta)} \mid N \right) \right] \leq e^{-cn\varepsilon_{n,\beta}},$$

and if $n^\rho \lesssim K_n = o(n/\log n)$ for some $\rho > 0$, then

$$\sup_{f_0 \in \text{PC}(K_n, R)} E_{f_0} \left[\Pi \left(f : \|f - f_0\|_1 \geq M \frac{K_n}{n} \log n \mid N \right) \right] \leq n^{-cK_n}.$$

The proof is a direct consequence of Theorem 2.3, Corollary 2.6 (ii), Lemma 5.1, and Lemma 5.2. In all cases the rate is expected to be optimal up to the $\log n$ factor.

5.2 Subordinators

A natural class of priors for monotone (non-decreasing) functions in the Skorokhod space $D[0, 1]$ are subordinators. A subordinator $(X_t)_{t \in [0, 1]}$ is a process with monotone, right-continuous sample paths, $X_0 = 0$ and independent, stationary increments. Hence, a subordinator is a Lévy process. We consider only pure-jump subordinators, which are characterized by their characteristic function

$$\phi_t(u) = \mathbb{E}[e^{iuX_t}] = \exp \left(\int_{\mathbb{R}^+} (e^{iux} - 1) \nu(dx) \right), \quad t \geq 0,$$

where the Lévy measure ν is a σ -finite measure on \mathbb{R}^+ , satisfying $\int_{\mathbb{R}^+} (x \wedge 1) \nu(dx) < \infty$. Its intensity is $\lambda = \nu(\mathbb{R}^+) \in [0, \infty]$ and in the finite intensity case a subordinator is just a compound Poisson process of intensity λ with jump distribution $G = \nu/\lambda$.

Among the subordinators of infinite intensity prominent examples are the Gamma process with $X_t \sim \Gamma(ct, d)$ with $c, d > 0$ and Lévy (Lebesgue) density $\nu(x) = cx^{-1}e^{-dx}$ and the inverse Gaussian process with $X_t \sim IG(\delta t, \gamma)$ and Lévy density $\nu(x) = \frac{\delta}{\sqrt{2\pi}} x^{-3/2} e^{-\gamma^2 x/2}$,

see the monograph [30]. Both processes can be easily simulated. For instance, the inverse Gaussian process with $\gamma = 0$ and $\delta = 1$ is obtained as the inverse function of the running maximum $M_t = \max_{s \leq t} W_s$ of a Brownian motion W .

Subordinators as priors have been studied in the context of survival models by [16]. There the target of estimation is the cumulative hazard function, which can be estimated at the parametric rate $n^{-1/2}$. Even more generally, there are only few results on posterior contraction under shape constraints. An exception is [29] proving contraction rates for monotone densities in nonparametric density estimation using mixtures of Dirichlet processes as priors. Subordinators as priors for shape-constrained estimation problems in regression or density-type models do not seem to have been analyzed yet such that the results below can be of independent interest.

5.4 Proposition. *Let $f_0 \in D[0, 1]$ be non-decreasing with $f_0(0) \geq 0$, $f_0(1) \leq R$ and let $(X_t)_{t \in [0, 1]}$ be a subordinator with finite intensity λ and jump distribution $G = \nu/\lambda$. If $\varepsilon > 0$ satisfies $\varepsilon^{-1} \in \mathbb{N}$, $\varepsilon \leq \lambda^{-1}$ and G has a Lebesgue density g with*

$$g(x) \geq cpe^{-\rho x}, \quad x \geq \varepsilon, \quad \text{for some } \rho > 0, c \in (0, 1],$$

then we have the non-asymptotic one-sided small ball probability bound

$$P(\|X - f_0\|_{L^1} \leq (2R + 2)\varepsilon, X \leq f_0) \geq e^{-\lambda - \rho(R+2)} (c\rho\lambda\varepsilon^2)^{\varepsilon^{-1}}.$$

For $\lambda \asymp \rho \asymp \varepsilon^{-1}$ the right-hand side can be lower bounded by $\bar{c}\varepsilon^{-1}$ for some $\bar{c} > 0$.

We turn to the potentially infinite intensity case. If the Lévy measure ν has a Lebesgue density, we abuse notation and call this Lévy density $\nu(x)$.

5.5 Proposition. *Let $\varepsilon \in (0, 1/2)$ and assume that $f_0 \in D[0, 1]$ is non-decreasing with $f_0(0) \geq \varepsilon$ and $f_0(1) \leq R$. Let $(X_t)_{t \in [0, 1]}$ be a subordinator with Lévy density ν . If $\nu(x) \geq c_1\rho e^{-\rho x}$, $x \geq \varepsilon$, and $\nu(x) \leq C_2x^{-3/2}$, $x < \varepsilon$, for some fixed $c_1, C_2, \rho > 0$ and $\int_\varepsilon^\infty \nu(x)dx \leq \varepsilon^{-1}$, then*

$$P(\|X - f_0\|_{L^1} \leq (2R + 3)\varepsilon, X \leq f_0) \geq \varepsilon^{-c\varepsilon^{-1}},$$

where $c > 0$ only depends on c_1, C_2, ρ, R .

The assumption $f_0(0) \geq \varepsilon$ occurs because we need one-sided approximations from below and $X_0 = 0$ with $\lambda = \infty$ implies $X_h > 0$ a.s. for any $h > 0$. If we additionally randomize the starting value of the subordinator, this condition can be avoided, but to remain concise this is not pursued further. Now let us study the one-sided separation quantity for the class of monotone functions in L^1 -distance.

5.6 Proposition. *Let $n \geq 16$ and consider*

$$\Theta = \left\{ f : [0, 1] \rightarrow \mathbb{R} \text{ non-decreasing} \right\}.$$

Then for any non-decreasing $f_0 \in D[0, 1]$ with $f_0(1) - f_0(0) \leq R$,

$$S_{\lceil} \left(n, \left\{ f \in \Theta \mid \int_0^{1-4/\sqrt{n}} (f - f_0)_+ \geq \frac{4R+8}{\sqrt{n}} \right\}, f_0 \right) \leq e^{-\sqrt{n}}.$$

Note that there is no further restriction for Θ besides monotonicity. In particular, Θ is not a compact subset of $L^1([0, 1])$ and standard entropy bounds do not exist. In our result, an upper bound $f(1) \leq R$ for some known R and the adjusted prior $\tilde{X}_t = X_t \wedge R$ would allow to extent the contraction rate to L^1 -balls on the entire interval $[0, 1]$ because on $[1-4/\sqrt{n}, 1]$ the integrand $(f - f_0)_+$ can just be bounded by R . We prefer the more general formulation above because it is well in line with the fact that estimation for monotone functions at the right boundary is intrinsically difficult since the monotone MLE $\hat{f}^{MLE}(x)$ diverges to infinity as $x \uparrow 1$, cf. [28].

5.7 Theorem. *Consider a subordinator prior satisfying the conditions of Proposition 5.4 with fixed $\lambda, \rho > 0$, or the conditions in Proposition 5.5. Then there are positive constants M, c , such that*

$$E_{f_0} \left[\Pi \left(f : \int_0^{1-4/\sqrt{n}} |f - f_0| \geq M \left(\frac{\log n}{n} \right)^{1/2} \mid N \right) \right] \leq e^{-c\sqrt{n \log n}}.$$

If the conditions of Proposition 5.4 with $\lambda_n \asymp \rho_n \asymp \sqrt{n}$ hold, then there are $M, c > 0$ such that

$$E_{f_0} \left[\Pi \left(f : \int_0^{1-4/\sqrt{n}} |f - f_0| \geq \frac{M}{\sqrt{n}} \mid N \right) \right] \leq e^{-c\sqrt{n}}.$$

The proof of this result is an immediate consequence of Propositions 2.4, 2.5, 5.4, 5.5, and 5.6.

6 Bernstein-von Mises type theorems and frequentist coverage of credible sets

6.1 An asymptotic shape result for the posterior

Studying the Bayesian method from a frequentist perspective, we are ideally interested in characterizations of the properly rescaled limiting distribution of the posterior. For non-parametric and high-dimensional models some Bernstein-von Mises type theorems have

been established recently, cf. [21, 3, 4, 5, 25, 27]. In this section, we derive an expression for the asymptotic posterior distribution for a class of priors on the piecewise constant functions with growing number of pieces.

Given a positive integer K and (t_0, t_1, \dots, t_K) with $0 =: t_0 < t_1 < \dots < t_K := 1$, consider the space of piecewise constant functions with fixed jump times:

$$\text{PC}^*(K, (t_0, \dots, t_K), R) = \{f : f = \sum_{j=1}^K a_j \mathbf{1}_{[t_{j-1}, t_j)}, |a_j| \leq R\}.$$

The underlying parameter vector is $\mathbf{a} = (a_1, \dots, a_K) \in [-R, R]^K$. We are mainly interested in the nonparametric regime with $K = K_n \rightarrow \infty$. For convenience we omit the dependence on n and write t_j for the n -dependent jump points. As prior density on the vector \mathbf{a} consider

$$\pi(\mathbf{a}) = \prod_{j=1}^K g(a_j) \quad (6.1)$$

with a fixed Lebesgue density g . This might be viewed as a simplification of the CPP prior with given model dimension K and grid points $(t_j)_j$. The MLE over $(a_1, \dots, a_K) \in \Theta = \mathbb{R}^K$ is

$$\hat{f}^{\text{MLE}} = \sum_{j=1}^K \hat{a}_j \mathbf{1}_{[t_{j-1}, t_j)}, \quad \text{with } \hat{a}_j := \min_{i: X_i \in [t_{j-1}, t_j)} Y_i.$$

Write $f_0 = \sum_{j=1}^K a_j^0 \mathbf{1}_{[t_{j-1}, t_j)}$ for the true function. Under P_{f_0} , we have $\hat{a}_j - a_0^j \sim \text{Exp}(n(t_j - t_{j-1}))$ because

$$P_{f_0}(\hat{a}_j - a_0 > y) = P_{f_0}(N([t_{j-1}, t_j) \times [a_0, a_0 + y]) = 0) = e^{-n(t_j - t_{j-1})y}, \quad y \geq 0.$$

The main result in this section provides simple conditions under which the posterior can be approximated in total variation by the conditional distribution of

$$\hat{f}^{\text{MLE}} - \sum_{j=1}^K \eta_j \mathbf{1}_{[t_{j-1}, t_j)} = \sum_{j=1}^K (\hat{a}_j - \eta_j) \mathbf{1}_{[t_{j-1}, t_j)},$$

for independent $\eta_j \sim \text{Exp}(n(t_j - t_{j-1}))$ given the data $(X_i, Y_i)_i$.

The posterior $\Pi(\cdot|N)$ on the vector (a_1, \dots, a_K) is a measure on $(\mathbb{R}^K, \mathcal{B}(\mathbb{R}^K))$. Let Q^n be the distribution of $(\hat{a}_1 - \eta_1, \dots, \hat{a}_K - \eta_K)$ on $(\mathbb{R}^K, \mathcal{B}(\mathbb{R}^K))$ for given $(\hat{a}_1, \dots, \hat{a}_K)$ and denote by $\|P_1 - P_2\|_{\text{TV}}$ the total variation distance between the probability measures P_1 and P_2 .

6.1 Theorem. *Fix $R > 0$. Consider the prior (6.1) and assume that g is positive and β -Hölder continuous for some $0 < \beta \leq 1$. If*

$$n \frac{\inf_{j=1, \dots, K_n} |t_j - t_{j-1}|}{K_n^{1/\beta} \log(K_n)} \rightarrow \infty, \quad (6.2)$$

then

$$\sup_{f_0 \in \text{PC}^*(K_n, (t_0, \dots, t_{K_n}), R)} E_{f_0} [\|\Pi(\cdot|N) - Q^n\|_{\text{TV}}] \rightarrow 0.$$

In contrast to the parametric Bernstein-von Mises theorem (cf. [33], Theorem 10.1), Theorem 6.1 also assumes Hölder smoothness on the marginal prior densities g . Together with the condition (6.2) this assures the right amount of smoothness of the full prior $\pi(\mathbf{a}) = \prod_{j=1}^{K_n} g(a_j)$. Indeed, since K_n might tend to infinity, it might otherwise happen that π becomes very rough for large n . The maximal speed at which K_n can tend to infinity in Theorem 6.1 depends on the Hölder index β and the rate at which the minimal grid length $t_j - t_{j-1}$ decreases. If the t_j are on a regular grid in the sense that $\inf_j (t_j - t_{j-1}) \gtrsim 1/K_n$, then

$$K_n = o\left(\frac{n}{\log n}\right)^{\beta/(\beta+1)}$$

suffices. For $\beta = 1$ we can allow K_n to be almost $n^{1/2}$. This should be compared to the Bernstein-von Mises phenomenon for increasing parameter dimension which requires a number of parameters smaller than $n^{1/3}$, cf. [25]. It is also worth mentioning [13] which establishes a weak Bernstein-von Mises theorem in the nonparametric regression model for a similar class of priors.

While the MLE overshoots each true parameters a_0^j by an exponential distribution with parameter $n(t_j - t_{j-1})$, asymptotically the posterior distribution "corrects" for that bias by subtracting independent η_j with the same distribution. This is the reason why Bayesian methods are advocated for related parametric boundary estimation problems in the frequentist literature, see e.g. the discussion in [7]. The Bernstein-von Mises result also shows that the prior is washed out in the limit. In the special case $K_n = 1$ the true function $f_0 = a_0^1$ is a constant and the corresponding likelihood is proportional to $e^{na_1} \mathbf{1}(a_1 \leq \min_i Y_i)$. The same likelihood is obtained in the model, where we observe n i.i.d. copies of $Y = a_1 + \varepsilon$ with $\varepsilon \sim \text{Exp}(1)$. This establishes the equivalence between Theorem 6.1 and Theorem 1.1 in [17] for $K_n = 1$.

6.2 A specific semi-parametric Bernstein-von Mises result

We study the Bernstein-von Mises phenomenon and frequentist coverage of credible sets for the functional $\vartheta = \int f$, which is the prototype of a smooth linear functional of f . For the class of piecewise constant functions, the MLE is $\hat{\vartheta}^{\text{MLE}} = \int \hat{f}^{\text{MLE}}$. By the explicit law of $\hat{\vartheta}^{\text{MLE}}$, we can derive $\hat{\vartheta}^{\text{MLE}} - \vartheta = K_n/n + O_P(\sqrt{K_n}/n)$. The MLE has thus rate of

convergence K_n/n whereas the bias corrected estimator $\hat{\vartheta} = \int \hat{f}^{\text{MLE}} - K_n/n$ attains the faster rate $O_P(\sqrt{K_n}/n)$.

The bias correction can also be derived by the general estimation theory for functionals as developed in [28]. For the integral $\vartheta = \int f$ the bias correction term is the number of points on the MLE divided by n . Since there are almost surely K_n points on the MLE for the parameter space $\text{PC}^*(K_n, (t_0, \dots, t_{K_n}), R)$, this provides us with another explanation of the bias correction term K_n/n .

6.2 Corollary. *Consider the prior (6.1) and work under the assumptions of Theorem 6.1. Denote by $B \mapsto \Pi(\vartheta \in B|N)$ the marginal posterior of the integral $\vartheta := \int f$. If $K_n \rightarrow \infty$, then*

$$\sup_{f_0 \in \text{PC}^*(K_n, (t_0, \dots, t_{K_n}), R)} E_{f_0} \left[\left\| \Pi(\vartheta \in \cdot | N) - \mathcal{N}\left(\hat{\vartheta}^{\text{MLE}} - \frac{K_n}{n}, \frac{K_n}{n^2}\right) \right\|_{\text{TV}} \right] \rightarrow 0.$$

The asymptotic $(1 - \alpha)$ -credible interval

$$I(\alpha) = \left[\hat{\vartheta}^{\text{MLE}} - \frac{K_n}{n} + \frac{\sqrt{K_n}}{n} \Phi^{-1}(\alpha/2), \hat{\vartheta}^{\text{MLE}} - \frac{K_n}{n} + \frac{\sqrt{K_n}}{n} \Phi^{-1}(1 - \alpha/2) \right] \quad (6.3)$$

is moreover also an honest asymptotic confidence set in the sense that

$$\inf_{f_0 \in \text{PC}^*(K_n, (t_0, \dots, t_{K_n}), R)} P_{f_0} \left(\int f_0 \in I(\alpha) \right) \rightarrow 1 - \alpha.$$

One of the interesting consequences of this result is that asymptotically for $K_n \rightarrow \infty$ the credible set does not contain the MLE $\hat{\vartheta}^{\text{MLE}} = \int \hat{f}^{\text{MLE}}$. Hence, the posterior distribution automatically corrects for the bias and is not misguided by the high values of the likelihood around $\hat{\vartheta}^{\text{MLE}}$.

6.3 A negative result on frequentist coverage of credible sets under model misspecification

We have shown that credible sets are asymptotic confidence sets for priors on the space of piecewise constant functions, provided the true function f_0 is also piecewise constant. In the frequentist estimation theory it is known that for Lipschitz-continuous functions f_0 a bias-corrected MLE over piecewise constant functions at jump points $t_j = j/K_n$ remains rate-optimal if $K_n \asymp n^{1/2}$, cf. the block-wise estimator in [28]. In the same spirit, the nonparametric Bayes result in Section 5 establishes good posterior contraction rates for Lipschitz functions given a CPP prior generating piecewise constant functions. As we shall see here, the automatic bias correction by a Bayes method may fail in the case of model misspecification. A consequence is that credible sets may have asymptotically no frequentist

coverage at all. We consider the same piecewise-constant prior as in the previous subsection with jump locations $t_j = j/K_n$ and study the limiting shape of the posterior for data generated by a piecewise linear function

$$f_0(x) = x + \sum_{j=1}^{K_n} a_j \mathbf{1}\left(x \in \left[\frac{j-1}{K_n}, \frac{j}{K_n}\right)\right). \quad (6.4)$$

We shall see that for a range of sequences K_n , in particular for $K_n \asymp n^{1/2}$, the credible set does not cover ϑ_0 . What happens is that the credible sets are still of the form (6.3), but the bias correction term K_n/n is not precise enough in this case. The next result explains this phenomenon in more detail showing that a confidence interval of the form $\hat{\vartheta}^{\text{MLE}} - K_n/n \pm c\sqrt{K_n}/n$ will not cover $\int f_0$ if $\sqrt{n} \leq K_n = o(n^{4/7})$.

6.3 Proposition. *Consider data generated by an f_0 of the form (6.4). Then the MLE taken over the class of piecewise constant functions*

$$\hat{f}^{\text{MLE}} = \sum_{j=1}^{K_n} \hat{a}_j^{\text{MLE}} \mathbf{1}_{[(j-1)/K_n, j/K_n)}$$

with $\hat{a}_j^{\text{MLE}} = \min_{i: X_i \in [(j-1)/K_n, j/K_n)} Y_i$ can be written in distribution as

$$\hat{a}_j = a_j^0 + \frac{j-1 + V_{jn}}{K_n},$$

where $(V_{jn})_j$ is i.i.d. with distribution defined by

$$P_{f_0}(V_{jn} \geq y) = \exp\left(-\frac{n}{2K_n^2}(y \wedge 1)^2 - \frac{n}{K_n^2}(y-1)_+\right), \quad y \geq 0.$$

Moreover, for $K_n \geq \sqrt{n}$ we have $\text{Var}_{f_0}(\hat{\vartheta}^{\text{MLE}})^{1/2} \lesssim \sqrt{K_n}/n$ and

$$E_{f_0}\left[\hat{\vartheta}^{\text{MLE}} - \frac{K_n}{n}\right] \leq \int f_0 - \frac{n}{2^7 K_n^3}.$$

As part of our negative result, we need to show that there is a frequentist method which does equally well for piecewise constant functions, but is also able to return converging confidence sets if the true function is of the form (6.4). Consider the space of piecewise 1-Lipschitz functions

$$\text{Lip}_{K_n} = \left\{ f : |f(x) - f(y)| \leq |x - y|, \forall x, y \in \left[\frac{j-1}{K_n}, \frac{j}{K_n}\right], \forall j = 1, \dots, K_n \right\}$$

and notice that this space contains all piecewise constant functions as well as all functions of the form (6.4).

6.4 Lemma. *Let $K_n \geq 1$, then for $0 < \alpha < 1$ there exists a frequentist confidence interval $C(\alpha)$ such that*

$$\inf_{f_0 \in \text{Lip}_{K_n}} P_{f_0} \left(\int f_0(x) dx \in C(\alpha) \right) \geq 1 - \alpha$$

and

$$\sup_{f_0 \in \text{Lip}_{K_n}} \text{length}(C(\alpha)) \lesssim \frac{\sqrt{K_n}}{n} + \frac{1}{\sqrt{K_n n}}.$$

To complete the argument, let us study the frequentist coverage of the credible sets. We consider a uniform prior on the function values to allow for a wider range of K_n as in Theorem 6.1.

6.5 Theorem. *Let $f_0(x) = x$ such that $\vartheta_0 = \int f_0(x) dx = 1/2$. Consider the prior (6.1) with $t_j = j/K_n$, $K_n \leq n/\log n$, $K_n \rightarrow \infty$, and $g(a_i) = (2R)^{-1} \mathbf{1}_{[-R, R]}$ for fixed $R > 3$. If $I(\alpha)$ is as in (6.3), then for $0 < \alpha < 1$*

$$E_{f_0} [\|\Pi(\cdot|N) - Q^n\|_{\text{TV}}] \rightarrow 0 \text{ and } E_{f_0} [\Pi(\vartheta \in I(\alpha) | N)] \rightarrow 1 - \alpha.$$

On the other hand, if $K_n = o(n^{4/7})$ and $a_n = 2^{-8}(nK_n^{-3/2} \wedge n^2 K_n^{-7/2})$, then

$$P_{f_0} \left(\vartheta_0 \leq \hat{\vartheta}^{\text{MLE}} - \frac{K_n}{n} + \frac{\sqrt{K_n}}{n} a_n \right) \rightarrow 0 \quad (6.5)$$

and in particular $I(\alpha)$ has asymptotically no frequentist coverage:

$$P_{f_0} (\vartheta_0 \in I(\alpha)) \rightarrow 0.$$

In parametric models a similar phenomenon has been observed in the case of model misspecification, cf. [18]. For nonparametric models, it is sometimes possible to take a ball that covers $1 - \alpha$ of the posterior and to show that enlarging the radius of the ball by a constant results in frequentist coverage tending to one, cf. [32]. The result above implies that in order to achieve frequentist coverage the radius needs to be multiplied by a sequence that tends to infinity. If $K_n \asymp \sqrt{n}$ the blow-up factor needs to be at least of the order $n^{1/4}$.

If one considers the Gaussian white noise model with the same prior, then it is not difficult to see that even in the misspecified model, credible sets would be asymptotic confidence sets. The difference is that in this model the average of the observations on each block is a sufficient statistic, while in the support boundary detection model the sufficient statistics are the blockwise minima. Taking local averages does not induce a bias if the true function was piecewise linear.

Acknowledgments

We thank Marc Hoffmann, Richard Nickl and the members of the Bayes Club in Amsterdam for helpful comments. Financial support by the DFG through research unit FOR 1735 *Structural Inference in Statistics: Adaptation and Efficiency* is gratefully acknowledged. In addition, the second author was partially supported by an NWO TOP grant.

7 Appendix

Proof of Lemma 2.1. The general change of measure formula for two Poisson point processes on \mathcal{X} with finite intensity measures $\Lambda_1 \ll \Lambda_2$ is given by

$$\frac{dQ_{\Lambda_1}}{dQ_{\Lambda_2}}(N) = \exp \left(\int_{\mathcal{X}} \log \frac{d\Lambda_1}{d\Lambda_2}(x) dN(x) - \Lambda_1(\mathcal{X}) + \Lambda_2(\mathcal{X}) \right), \quad (7.1)$$

where $\log 0 := -\infty$, $\exp(-\infty) := 0$, cf. [20], Theorem 1.3. Notice that P_f and P_g have infinite intensity. We therefore apply the following decomposition first. For two functions $h_1 \leq h_2$ denote by P_{h_1, h_2} the distribution of a PPP on $[0, 1] \times \mathbb{R}$ with intensity $\lambda_{h_1, h_2}(x, y) := n \mathbf{1}(h_1(x) \leq y \leq h_2(x))$. Given a PPP with distribution P_f we can decompose it in two independent PPPs with distribution $P_{f, f \vee g}$ and $P_{f \vee g}$, that is, $P_f = P_{f, f \vee g} \otimes P_{f \vee g}$. Similarly, $P_g = P_{g, f \vee g} \otimes P_{f \vee g}$. The PPPs $P_{f, f \vee g}$ and $P_{g, f \vee g}$ have finite intensity and we can apply (7.1) with $\mathcal{X} = [0, 1] \times \mathbb{R}$, $\Lambda_1(A) = \int_A \lambda_{f, f \vee g}(x, y) dx dy$, and $\Lambda_2(A) = \int_A \lambda_{g, f \vee g}(x, y) dx dy$. With $N_{< f \vee g} := \sum_{i: Y_i < (f \vee g)(X_i)} \delta_{(X_i, Y_i)}$, this gives

$$\begin{aligned} \frac{dP_f}{dP_g}(N) &= \frac{dP_{f, f \vee g}}{dP_{g, f \vee g}}(N_{< f \vee g}) \\ &= \exp \left(\sum_{i: Y_i < (f \vee g)(X_i)} \frac{n \mathbf{1}(Y_i \geq f(X_i))}{n \mathbf{1}(Y_i \geq g(X_i))} - n \int (f \vee g - f) + n \int (f \vee g - g) \right) \\ &= e^{n \int (f - g)} \mathbf{1}(\forall i : Y_i \geq f(X_i)). \end{aligned}$$

□

Proof of Lemma 2.2. Let $f_0 \in L^1([0, 1])$ be fixed. Consider a PPP with a positive intensity λ^* satisfying

$$\lambda^*(x, y) = n \mathbf{1}(y \geq f_0(x)), \quad \Lambda^*(\{y < f_0(x)\}) = \int_0^1 \int_{-\infty}^{f_0(x)} \lambda^*(x, y) dy dx < \infty,$$

and denote by P^* its distribution. For any $f \in L^1([0, 1])$ we can decompose the process into two independent PPP with intensities $\lambda_{f, 1}^*(x, y) = \lambda^*(x, y) \mathbf{1}(y < f \vee f_0(x))$

and $\lambda_{f \vee f_0}(x, y) = n \mathbf{1}(y \geq f \vee f_0(x))$. If $P_{f,1}^*$ and $P_{f \vee f_0}$ denote the corresponding laws, then we can decompose $P^* = P_{f,1}^* \otimes P_{f \vee f_0}$. Similarly, we can decompose the distribution P_f via $P_f = P_{f,1} \otimes P_{f \vee f_0}$, where $P_{f,1}$ denotes the PPP distribution with intensity $\lambda_{f,1}(x, y) = n \mathbf{1}(f(x) \leq y \leq f_0(x))$. Finally, for a PPP N we write $N = N_{<f \vee f_0} + N_{\geq f \vee f_0}$ with $N_{<f \vee f_0} = \sum_{i: Y_i < f \vee f_0(X_i)} \delta_{(X_i, Y_i)}$ and $N_{\geq f \vee f_0} = \sum_{i: Y_i \geq f \vee f_0(X_i)} \delta_{(X_i, Y_i)}$. Then, using (7.1),

$$\begin{aligned} \frac{dP_f}{dP^*}(N) &= \frac{dP_{f,1}}{dP_{f,1}^*}(N_{<f \vee f_0}) \\ &= e^{-n \int (f_0 - f)_+ + \Lambda^*({y < f_0(x)}) + n \int (f - f_0)_+} \prod_{i: Y_i < f \vee f_0(X_i)} \frac{\lambda_{f,1}}{\lambda_{f,1}^*}(X_i, Y_i) \\ &= e^{n \int (f - f_0)_+ + \Lambda^*({y < f_0(x)})} \mathbf{1}(\forall i : f(X_i) \leq Y_i) \prod_{i: Y_i < f \vee f_0(X_i)} \frac{n}{\lambda^*(X_i, Y_i)}, \end{aligned}$$

where products over empty index sets are set to one. Now, notice that under P_{f_0} we have $Y_i \geq f_0(X_i)$ and thus $\lambda^*(X_i, Y_i) = n$ a.s. such that

$$\Pi(B|N) = \frac{\int_B \frac{dP_f}{dP^*}(N) d\Pi(f)}{\int \frac{dP_f}{dP^*}(N) d\Pi(f)} = \frac{\int_B e^{n \int f} \mathbf{1}(\forall i : f(X_i) \leq Y_i) d\Pi(f)}{\int e^{n \int f} \mathbf{1}(\forall i : f(X_i) \leq Y_i) d\Pi(f)} \quad P_{f_0}\text{-a.s.}$$

Under P_{f_0} we have $\mathbf{1}(\forall i : f(X_i) \leq Y_i) = \mathbf{1}(\forall i : f \vee f_0(X_i) \leq Y_i)$ a.s. and (2.1) yields

$$H(f) := e^{n \int (f - f_0)_+} \mathbf{1}(\forall i : f \vee f_0(X_i) \leq Y_i) = e^{-n \int (f_0 - f)_+} \frac{dP_{f \vee f_0}}{dP_{f_0}}(N), \quad (7.2)$$

which completes the proof. \square

Proof of Proposition 2.4. Consider $H(f)$ from (7.2). By Lemma 2.2, under P_{f_0}

$$\Pi(B|N) = \frac{\int_B H(f) d\Pi(f)}{\int H(f) d\Pi(f)} \leq e^{A n \varepsilon_n} \frac{\int_B H(f) d\Pi(f)}{\Pi(f : \|f - f_0\|_1 \leq A \varepsilon_n \text{ \& } f \leq f_0)}, \quad (7.3)$$

where we used $\mathbf{1}(\forall i : f(X_i) \leq Y_i) = 1$ P_{f_0} -a.s. for all $f \leq f_0$. With $B := \{f : \int (f_0 - f)_+ \geq (1 + A + C) \varepsilon_n\}$ and $E_{f_0}[H(f)] = e^{-n \int (f_0 - f)_+} \leq e^{-(1 + A + C) \varepsilon_n}$ for $f \in B$, we obtain $E_{f_0}[\Pi(B|N)] \leq e^{-n \varepsilon_n}$. \square

Proof of Proposition 2.5. For functions $(\ell_j)_{j \in J}$ eligible in the definition of $S[(n, B, f_0)]$ consider the test $\phi_n = \mathbf{1}(\exists j \forall i : \ell_j(X_i) \leq Y_i)$. This test satisfies under the hypothesis f_0

$$P_{f_0}(\phi_n = 1) \leq \sum_{j \in J} P_{f_0}(\forall i : \ell_j(X_i) \leq Y_i) = \sum_{j \in J} e^{-n \int (\ell_j - f_0)_+}.$$

By assumption and σ -continuity of Π , there exist $R > 0$ and $\delta > 0$ such that $\Pi(f : \int f \geq -R, f \leq f_0) \geq \delta$. Thus, we use formula (7.3) and bound the posterior by

$$\Pi(B|N) \leq \phi_n + \frac{\int_B H(f)(1 - \phi_n)d\Pi(f)}{\int H(f)d\Pi(f)} \leq \phi_n + \delta^{-1}e^{nR+n\int f_0} \int_B H(f)(1 - \phi_n)d\Pi(f).$$

Since for $f \in B$ there is an $\ell_j \leq f$, we infer

$$H(f)(1 - \phi_n) = e^{n\int(f-f_0)} \mathbf{1}(\forall i : f(X_i) \leq Y_i) \mathbf{1}(\forall j \exists i : \ell_j(X_i) > Y_i) = 0.$$

Therefore,

$$E_{f_0}[\Pi(B|N)] \leq E_{f_0}[\phi_n] \leq \sum_{j \in J} e^{-n\int(\ell_j - f_0)_+}$$

and the claim follows by taking the infimum over all possible (ℓ_j) . \square

Proof of Theorem 2.3. By Proposition 2.4 and Proposition 2.5 it remains to show $E_{f_0}[\Pi(\Theta_n^c|N)] \leq C''e^{-n\varepsilon_n}$. By (7.3), $E_{f_0}[H(f)] \leq 1$ and condition (i) and (iii),

$$E_{f_0}[\Pi(\Theta_n^c|N)] \leq \frac{e^{nA\varepsilon_n}\Pi(\Theta_n^c)}{\Pi(f : \|f - f_0\|_1 \leq A\varepsilon_n, f \leq f_0)} \leq C''e^{-n\varepsilon_n}$$

follows, which is the claim. \square

Proof of Lemma 2.7. The key observation is that we can restrict the posterior to $f \leq \hat{f}^{\text{MLE}}$ since otherwise the likelihood is zero. To see this, note that $\forall i : f(X_i) \leq Y_i$ implies $f \leq \hat{f}^{\text{MLE}}$ because otherwise $f \vee \hat{f}^{\text{MLE}} \in \Theta_n$ would have a larger likelihood than \hat{f}^{MLE} . Observe that then $\int(f - f_0)_+ \leq \int(\hat{f}^{\text{MLE}} - f_0)_+$ such that

$$\begin{aligned} E_{f_0}\left[\Pi\left(f \in \Theta_n : \int(f - f_0)_+ > \varepsilon_n \middle| N\right)\right] &\leq E_{f_0}\left[\Pi\left(f \in \Theta_n : \int(\hat{f}^{\text{MLE}} - f_0)_+ > \varepsilon_n \middle| N\right)\right] \\ &= P_{f_0}\left(\int(f^{\text{MLE}} - f_0)_+ > \varepsilon_n\right), \end{aligned}$$

where the last equality holds because $\{\int(\hat{f}^{\text{MLE}} - f_0)_+ > \varepsilon_n\}$ is independent of f . \square

8 Proofs for Section 3

We state Theorem 2.1 of [34] in a slightly more general form.

8.1 Theorem (Theorem 2.1 of [34]). *Let X be a Borel-measurable, zero-mean Gaussian random element in the Banach space $(\mathbb{B}, \|\cdot\|)$ with RKHS $(\mathbb{H}, \|\cdot\|_{\mathbb{H}})$ and let f be contained in the closure of \mathbb{H} in \mathbb{B} . For any $\varepsilon_n > 0$ and $\gamma_n \geq 1$, satisfying*

$$\inf_{h: \|h-f\| \leq \varepsilon_n} \|h\|_{\mathbb{H}}^2 - \log P(\|X\| \leq \varepsilon_n) \leq \gamma_n$$

and for any $C_ \geq 1$, there exists a Borel set $B_n \subset \mathbb{B}$ such that*

$$\log N(3\varepsilon_n, B_n, \|\cdot\|) \leq 6C_*\gamma_n, \quad \mathbb{P}(X \notin B_n) \leq e^{-C_*\gamma_n}, \quad \text{and} \quad \mathbb{P}(\|X - f\| \leq 2\varepsilon_n) \geq e^{-\gamma_n}.$$

Proof. Replace $n\varepsilon_n^2$ in the proof of Theorem 2.1 of [34] by γ_n ; in particular $M_n := -2\Phi^{-1}(e^{-C_*\gamma_n})$. For the final argument of the proof observe that $e^{-C_*\gamma_n} < 1/2$ due to $C_*\gamma_n \geq 1$. \square

Proof of Theorem 3.1. We apply Theorem 8.1 with $(\mathbb{B}, \|\cdot\|) = (\mathcal{C}[0, 1], \|\cdot\|_{\infty})$, $\gamma_n = n\varepsilon_n$ and $C_* = 6$. This shows that there exists Θ_n such that $\log N(3\varepsilon_n, \Theta_n, \|\cdot\|_{\infty}) \leq 36n\varepsilon_n$, $\Pi(\Theta_n^c) \leq e^{-6n\varepsilon_n}$ and $\Pi(f : \|f + 2\varepsilon_n - f_0\|_{\infty} \leq 2\varepsilon_n) \geq e^{-n\varepsilon_n}$. Use (2.3) and (iii) in Corollary 2.6. Thus, the assumptions of Theorem 2.3 are satisfied with $A = 4, C = C'' = 1, C' = 36$. \square

9 Proofs for Section 4

Proof of Lemma 4.1. Write $h = \sum_{j,k} h_{j,k} \psi_{j,k}$. Since ψ is a compactly supported wavelet, $\|X - h\|_{\infty} \leq C \sum_j 2^{j/2} \max_k |d_{j,k} \xi_{j,k} - h_{j,k}|$ for a sufficiently large constant C . By assumption ψ is moreover s -regular and $h \in \mathcal{C}^{\beta}(R)$ with $\beta \leq s$. Using Theorem 4.4 in [8], we can find constants $0 < q < Q < \infty$ such that $q2^{-\frac{j}{2}(2\alpha+1)} \leq |d_{j,k}| \leq Q2^{-\frac{j}{2}(2\alpha+1)}$ and $|h_{j,k}| \leq Q2^{-\frac{j}{2}(2\beta+1)}$ and obtain for any J ,

$$\|X - h\|_{\infty} \leq CQ \left(\sum_{j \leq J} 2^{-j\alpha} \max_k |\xi_{j,k} - h_{j,k}/d_{j,k}| + \sum_{j > J} 2^{-j\alpha} \max_k |\xi_{j,k}| + \sum_{j > J} 2^{-j\beta} \right).$$

Let J_* be the smallest integer such that

$$CQ \left(2^{-J_*\alpha} \sum_{r=0}^{J_*} 2^{-\alpha r} + L^{\alpha/(1+\delta)} 2^{-J_*\alpha} \sum_{r \geq 1} 2^{-r\alpha\delta/(2+\delta)} + 2^{-J_*\beta} \sum_{r \geq 0} 2^{-r\beta} \right) \leq \varepsilon,$$

which yields $2^{J_*} \asymp \varepsilon^{-1/(\alpha \wedge \beta)}$ as $\varepsilon \rightarrow 0$. Introduce the events $G_{\leq} = \{|\xi_{j,k} - h_{j,k}/d_{j,k}| \leq 2^{(j-J_*)2\alpha} \text{ for } j \leq J_*\}$ and $G_{>} = \{|\xi_{j,k}| \leq L^{\alpha/(1+\delta)} 2^{(j-J_*)\alpha/(1+\delta/2)} \text{ for } j > J_*\}$ then, thanks to the choice of J_* , on $G_{\leq} \cap G_{>}$ we have $\sup_{t \in [0,1]} |X_t - h(t)| \leq \varepsilon$. Thus,

$$\mathbb{P}(\|X - h\|_{\infty} \leq \varepsilon) \tag{9.1}$$

$$\geq \prod_{j \leq J_*, k} \mathbb{P}(|\xi_{j,k} - h_{j,k}/d_{j,k}| \leq 2^{(j-J_*)2\alpha}) \prod_{j > J_*, k} \mathbb{P}(|\xi_{j,k}| \leq L^{\alpha/(1+\delta)} 2^{(j-J_*)\alpha/(1+\delta/2)}).$$

On the event G_{\leq} we also have for $j \leq J_*$

$$|\xi_{j,k}| \leq 2^{(j-J_*)2\alpha} + |h_{j,k}/d_{j,k}| \leq 2^{(j-J_*)2\alpha} + q^{-1}Q2^{j(\alpha-\beta)} \leq R'2^{J_*(\alpha-\beta)+}$$

with a sufficiently large constant $R' \geq 1$. Since the random variables $\xi_{j,k}$ are symmetric and have a unimodal density, we have $f_{\xi}(x) \leq 1/2$ for $x \geq 1$ as well as $\mathbb{P}(|\xi_{j,k} - h_{j,k}/d_{j,k}| \leq 2^{(j-J_*)2\alpha}) \geq 2^{(j-J_*)2\alpha} f_{\xi}(R'2^{J_*(\alpha-\beta)+})$. On the j -th resolution level there are at most $A2^j$ wavelet coefficients with A some positive constant. The first product in (9.1) can therefore be bounded from below by

$$\begin{aligned} \prod_{j \leq J_*} f_{\xi}(R'2^{J_*(\alpha-\beta)+})^{A2^j} (2^{(j-J_*)2\alpha})^{A2^j} &\geq f_{\xi}(R'2^{J_*(\alpha-\beta)+})^{A2^{J_*}} \prod_{k=1}^{J_*} 2^{-2\alpha k 2^{-k} A2^{J_*}} \\ &\geq f_{\xi}(R'2^{J_*(\alpha-\beta)+})^{A2^{J_*}} K^{-2^{J_*}} \end{aligned} \quad (9.2)$$

for a sufficiently large constant K . To find a lower bound of the second product in (9.1), observe that

$$\begin{aligned} \mathbb{P}(|\xi_{j,k}| \leq L^{\alpha/(1+\delta)} 2^{(j-J_*)\alpha/(1+\delta/2)}) &= 1 - \mathbb{P}(|\xi_{j,k}|^{(1+\delta)/\alpha} > L2^{(j-J_*)(1+\delta)/(1+\delta/2)}) \\ &\geq 1 - 2^{(J_*-j)(1+\delta)/(1+\delta/2)}. \end{aligned}$$

For any fixed $j > J_*$, using $(1+\delta)/(1+\delta/2) = 1 + \delta/(2+\delta)$ and the elementary inequality $1 - y \geq e^{-2y}$ for $0 \leq y \leq 1/2$,

$$\begin{aligned} \prod_k \mathbb{P}(|\xi_{j,k}| \leq L2^{(j-J_*)(1+\delta)/(1+\delta/2)}) &\geq (1 - 2^{(J_*-j)(1+\delta)/(1+\delta/2)})^{A2^j} \\ &\geq \exp(-A2^{J_*+1} 2^{(J_*-j)\delta/(2+\delta)}). \end{aligned}$$

This implies that the product $\prod_{j > J_*, k} \mathbb{P}(|\xi_{j,k}| \leq L^{\alpha/(1+\delta)} 2^{(j-J_*)\alpha/(1+\delta/2)})$ can be bounded from below by

$$\prod_{j > J_*} \exp(-A2^{J_*+1} 2^{(J_*-j)\delta/(2+\delta)}) = \exp(-A2^{J_*+1} \sum_{k \geq 1} 2^{-k\delta/(2+\delta)}) \geq \exp(-R''2^{J_*}) \quad (9.3)$$

for a sufficiently large constant R'' . Recall that $2^{J_*} \asymp \varepsilon^{-1/(\alpha \wedge \beta)}$. Because of $f_{\xi}(x) \leq 1/2$ for $|x| \geq 1$ we have $K^{-2^{J_*}} \exp(-R''2^{J_*}) \geq f_{\xi}(K'\varepsilon^{-(\alpha-\beta)/\beta})^{K'2^{J_*}}$ for a sufficiently large constant K' . The result follows therefore from (9.1), (9.2), and (9.3). \square

9.1 Lemma. *Suppose that for some $\gamma > 0$, $f_{\xi}(x) \leq \gamma^{-1}e^{-\gamma|x|}$ for all $x \in \mathbb{R}$. Let $\xi_1, \dots, \xi_m \sim f_{\xi}$, independently. Then,*

$$\mathbb{P}\left(\frac{1}{m} \sum_{j=1}^m |\xi_j| \geq 2\gamma^{-1}(t + \log(4/\gamma^2))\right) \leq e^{-tm}.$$

Proof. Set $A(\gamma, t) := 2\gamma^{-1}(t + \log(4/\gamma^2))$. By assumption, $f_{|\xi|}(x) \leq 2\gamma^{-1}e^{-\gamma x}$ holds for all $x \geq 0$ such that $E[e^{\gamma|\xi|/2}] \leq 4/\gamma^2$. Therefore, we deduce by Markov's inequality

$$\begin{aligned} \mathbb{P}\left(\sum_{j=1}^m |\xi_j| \geq A(\gamma, t)m\right) &= \mathbb{P}\left(\exp\left(\frac{\gamma}{2} \sum_{j=1}^m |\xi_j|\right) \geq \exp\left(\frac{\gamma}{2} A(\gamma, t)m\right)\right) \\ &\leq \exp\left(m \log\left(\frac{4}{\gamma^2}\right) - \frac{\gamma}{2} A(\gamma, t)m\right) = e^{-tm}. \end{aligned}$$

□

Proof of Theorem 4.2. It is enough to prove the result for all $n \geq n_0$ with n_0 a fixed integer. We verify the conditions (i) – (iii) of Theorem 2.3.

(i): To check the first condition, pick J_n such that $\varepsilon_n^{-1/\alpha} \leq 2^{J_n} < 2\varepsilon_n^{-1/\alpha}$. For a constant K , which will be chosen later to be large enough, define

$$\Theta_n = \left\{ g = \sum_{j,k} \theta_{j,k} \psi_{j,k} : \sum_{j \leq J_n} 2^{\frac{j}{2}(2\alpha+1)} \sum_k |\theta_{j,k}| \leq K 2^{J_n}, \max_{j > J_n} 2^{\frac{j}{2}(2\alpha-1)} \sum_k |\theta_{j,k}| \leq K \right\}.$$

Denote by

$$B_{p,q}^s(M) := \left\{ g = \sum_{j,k} \theta_{j,k} \psi_{j,k} : \left(\sum_j 2^{qj(s+\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\theta_{j,k}|^p \right)^{q/p} \right)^{1/q} \leq M \right\}$$

the Besov $B_{p,q}^s$ -ball with radius M and apply the usual modifications for $p = \infty$ and $q = \infty$. To bound the bracketing entropy of Θ_n , observe that $\Theta_n \subseteq B_{1,1}^{\alpha+1}(K 2^{J_n}) + B_{1,\infty}^\alpha(K)$ where the sum is the elementwise addition. By Theorem 4.3.36 in [12], there exists a constant C such that $\log \mathcal{N}(\delta, B_{p,q}^s(M), \|\cdot\|_\infty) \leq C(M/\delta)^{1/s}$ if $s > 1/p$. Since $2^{J_n} \asymp \varepsilon_n^{-1/\alpha}$ and $\alpha > 1$, this gives

$$\begin{aligned} \log \mathcal{N}(C\varepsilon_n, \Theta_n, \|\cdot\|_\infty) &\leq \log \mathcal{N}(C\varepsilon_n/2, B_{1,1}^{\alpha+1}(K 2^{J_n}), \|\cdot\|_\infty) + \log \mathcal{N}(C\varepsilon_n/2, B_{1,\infty}^\alpha(K), \|\cdot\|_\infty) \\ &\lesssim (C\varepsilon_n)^{-1/\alpha}. \end{aligned}$$

Notice that $\varepsilon_n \gtrsim n^{-\alpha/(1+\alpha)}$. Making the constant C big enough, we therefore obtain $\mathcal{N}(C\varepsilon_n, \Theta_n, \|\cdot\|_\infty) \leq e^{Cn\varepsilon_n}$.

(ii): We apply Lemma 4.1 to (2.3). Since $\varepsilon_n \rightarrow 0$, $f_0 - \varepsilon_n \in \mathcal{C}^\beta(R+1)$ for sufficiently large n and

$$\Pi(f : \|f + \varepsilon_n - f_0\|_\infty \leq \varepsilon_n) \geq f_\xi(D\varepsilon_n^{-(\alpha-\beta)+/\beta})^{D\varepsilon_n^{-1/(\alpha \wedge \beta)}} \geq e^{-Cn\varepsilon_n}.$$

(iii): We bound $\Pi(\Theta_n^c)$. Recall that $X = \sum_{j,k} d_{j,k} \xi_{j,k} \psi_{j,k}$ and $|d_{j,k}| \leq Q 2^{-\frac{j}{2}(2\alpha+1)}$ for all j, k . Thus,

$$\Pi(\Theta_n^c) \leq \mathbb{P}\left(\sum_{j \leq J_n} 2^{\frac{j}{2}(2\alpha+1)} \sum_k |d_{j,k} \xi_{j,k}| \geq K 2^{J_n}\right) + \mathbb{P}\left(\max_{j > J_n} 2^{\frac{j}{2}(2\alpha-1)} \sum_k |d_{j,k} \xi_{j,k}| \geq K\right)$$

$$\leq \mathbb{P}\left(Q \sum_{j \leq J_n} \sum_k |\xi_{j,k}| \geq 2^{J_n} K\right) + \sum_{j > J_n} \mathbb{P}\left(Q 2^{-j} \sum_k |\xi_{j,k}| \geq K\right).$$

On the j -th resolution level there are of the order of 2^j wavelet coefficients. Recall that $2^{J_n} \asymp \varepsilon_n^{-1/\alpha}$ and $\varepsilon_n \gtrsim n^{-\alpha/(1+\alpha)}$. Lemma 9.1 shows now that for any constant c , $\Pi(\Theta_n^c) \leq e^{-cn\varepsilon_n}$ for sufficiently large K .

The assertion follows by Theorem 2.3. \square

Proof of Lemma 4.4. Since ψ is s -regular and $\beta \leq s$, $|h_{j,k}| \lesssim 2^{-\frac{j}{2}(2\alpha+1)}$. As ψ has compact support, there exists a constant C such that $\|X - h\|_\infty \leq C(\sum_{j \leq J} \max_k |\xi_{j,k} - h_{j,k}| + 2^{-J\beta})$. Let J^* be the smallest integer such that $C(\sum_{j \leq J} \varepsilon/(2JC) + 2^{-J\beta}) \leq \varepsilon$. Notice that $2^{J^*} \asymp \varepsilon^{-1/\beta}$ as $\varepsilon \rightarrow 0$ and

$$\mathbb{P}(\|X - h\|_\infty \leq \varepsilon) \geq \mathbb{P}(J = J^*) \prod_{j \leq J^*, k} \mathbb{P}(|\xi_{j,k} - h_{j,k}| \leq \varepsilon/(2J^*C)).$$

Let $c := \inf_{x: |x| \leq (2J^*C)^{-1} + \max_{j \leq J^*, k} |h_{j,k}|} f_\xi(x)$. By the assumptions on f_ξ , we can conclude that $c > 0$. Together with $\mathbb{P}(J = J^*) \propto 2^{-J^*}$ and the fact that on the j -th resolution level the number of wavelet coefficients is bounded by $A2^j$ for some $A > 0$, this shows that

$$\mathbb{P}(\|X - h\|_\infty \leq \varepsilon) \gtrsim 2^{-J^*} \left(\frac{c\varepsilon}{J^*}\right)^{A2^{J^*}}.$$

The result follows from $\varepsilon \asymp 2^{J^*}$. \square

Proof of Theorem 4.5. We verify the conditions (i) – (iii) of Theorem 2.3.

(i): To check the first condition, pick J_n such that $\varepsilon_n^{-1/\beta} \leq 2^{J_n} < 2\varepsilon_n^{-1/\beta}$. For a constant K that will be chosen later to be large enough, define

$$\Theta_n = \left\{ g = \sum_{j \leq J, k} \theta_{j,k} \psi_{j,k} : J \leq J_n, \sum_{j \leq J, k} |\theta_{j,k}| \leq K2^{J_n} \right\}.$$

As the wavelet has compact support, we have for $g = \sum_{j \leq J, k} \theta_{j,k} \psi_{j,k}$ and $h = \sum_{j \leq J, k} \theta'_{j,k} \psi_{j,k}$, the sup-norm bound $\|g - h\|_\infty \leq C'2^{J/2} \sum_{j \leq J, k} |\theta_{j,k} - \theta'_{j,k}|$ for some positive constant C' . From [12], Proposition 4.3.36,

$$\mathcal{N}(C\varepsilon_n, \Theta_n, \|\cdot\|_\infty) \leq \left(\frac{C''2^{3J_n/2}}{C\varepsilon_n}\right)^{2^{J_n}}.$$

Since $\varepsilon_n^{-1/\beta} \leq 2^{J_n} < 2\varepsilon_n^{-1/\beta}$ and $\varepsilon_n = (\log n/n)^{-\beta/(1+\beta)}$, we therefore obtain $\mathcal{N}(C\varepsilon_n, \Theta_n, \|\cdot\|_\infty) \leq e^{Cn\varepsilon_n/2}$ provided C is chosen sufficiently large. Now, (i) follows from Corollary 2.6.

(ii): Since $\varepsilon_n \rightarrow 0$, $f_0 - \varepsilon_n \in \mathcal{C}^\beta(R+1)$ for sufficiently large n . The result follows from applying Lemma 4.4 to (2.3) and

$$\Pi(f : \|f + \varepsilon_n - f_0\|_\infty \leq \varepsilon_n) \geq \varepsilon_n^{D\varepsilon_n^{-1/\beta}} \geq e^{-Cn\varepsilon_n},$$

for sufficiently large C .

(iii): Observe that $\Pi(\Theta_n^c) \leq \mathbb{P}(J > J_n) + \mathbb{P}(\sum_{j \leq J_n, k} |\xi_{j,k}| \geq K2^{J_n})$. The sum $\sum_{j \leq J_n, k}$ is over $A2^{J_n}$ wavelet coefficients. Recall that $2^{J_n} \asymp \varepsilon_n^{-1/\alpha}$ and $\varepsilon_n = (\log n/n)^{\beta/(1+\beta)}$. Lemma 9.1 shows now that for any constant c we obtain $\Pi(\Theta_n^c) \leq e^{-cn\varepsilon_n}$ for sufficiently large K .

The assertion follows from Theorem 2.3. \square

10 Proofs for Section 5

Proof of Lemma 5.1. Proof of (i) : Let δ be such that $\frac{1}{2}(\varepsilon/(2R))^{1/\beta} \leq \delta \leq (\varepsilon/(2R))^{1/\beta}$ and $N := 1/\delta$ is a positive integer. Let $r(0, \delta) := f(0)$ and $r(j, \delta) := f(j\delta) - f((j-1)\delta)$ for $j \geq 1$. Define the step function

$$h := \sum_{j=0}^{N-1} f(j\delta) \mathbf{1}_{[j\delta, (j+1)\delta)} = \sum_{j=0}^{N-1} r(j, \delta) \mathbf{1}_{[j\delta, 1]}$$

Now, $\delta \leq (\varepsilon/(2R))^{1/\beta}$ and $f \in \mathcal{C}^\beta(R)$ give $\|f - h\|_\infty \leq \varepsilon/2$. It is thus enough to prove $\mathbb{P}(\|X - h\|_\infty \leq \varepsilon/2) \geq \varepsilon^{-c\varepsilon^{-1/\beta}}$. By assumption, g is continuous and positive and therefore also $c_0 := \inf_{-2R-1 \leq x \leq 2R} g(x)$ is positive. Due to (5.2), $|r(j, \delta)| \leq 2R$, $e^{-\lambda}/\lambda \geq e^{-2\lambda}$, and $N! \leq N^N$,

$$\begin{aligned} & \mathbb{P}(\|X - h\|_\infty \leq \varepsilon/2) \\ & \geq \Pi\left(k = N-1, r(j, \delta) - \frac{\varepsilon\delta}{4} \leq a_j \leq r(j, \delta), t_j \in \left[j\delta, j\delta + \frac{\varepsilon\delta}{4}\right], j = 0, \dots, N-1\right) \\ & \geq e^{-\lambda} \frac{\lambda^{N-1}}{(N-1)!} \left(c_0 \frac{\varepsilon\delta}{4}\right)^N \left(\frac{\varepsilon\delta}{4}\right)^{N-1} \\ & \geq e^{-2\lambda} (1 \wedge \lambda)^{2(2R/\varepsilon)^{1/\beta}} \left(\frac{\varepsilon}{2R}\right)^{1/(\beta\delta)} \left(\sqrt{c_0} \frac{\varepsilon}{4}\right)^{2/\delta}. \end{aligned}$$

Proof of (ii) : Let $f = \sum_{j=1}^{K_n} a_j \mathbf{1}_{[t_{j-1}, t_j]}$ be an arbitrary function in $\text{PC}(K_n, R)$. Without loss of generality, we can assume that $R \geq 2$. Choose δ such that $\varepsilon/(4RK_n) \leq \delta \leq \varepsilon/(2RK_n)$ and $N := 1/\delta$ is a positive integer. Define $h = \sum_{j=1}^N \min_{x \in [(j-1)\delta, j\delta]} f(x) \mathbf{1}_{[(j-1)\delta, j\delta]}$. Obviously $h \leq f$ and $\|f - h\|_1 \leq K_n R \delta \leq \varepsilon/2$. We can then write $h = \sum_{j=0}^{K_n^*} b_j^* \mathbf{1}_{[t_j^*, 1]}$ with $K_n^* \leq K_n$, $|b_j^*| \leq 2R$, $0 = t_0^* < t_1^* < \dots < t_{K_n^*}^* < 1$ and t_j^* a multiple of δ (only incorporating points $j'\delta$

where h actually jumps). Let I_j denote the interval with endpoints t_j^* and $t_j^* + \delta \text{sign}(\Delta_j)/2$. Let $c_0 := \inf_{-2R-1 \leq x \leq 2R} g(x)$. Arguing as in (i), $c_0 > 0$ and

$$\begin{aligned} & \mathbb{P}(\|X - h\|_1 \leq \varepsilon, X \leq h) \\ & \geq \Pi\left(k = K_n^*, b_j^* - \delta/2 \leq b_j \leq b_j^*, t_k \in I_k, j = 0, \dots, K_n^*, k = 1, \dots, K_n^*\right) \\ & \geq e^{-\lambda} \frac{\lambda^{K_n^*}}{K_n^*!} \left(c_0 \frac{\delta}{2}\right)^{K_n^*+1} \left(\frac{\delta}{2}\right)^{K_n^*} \\ & \geq e^{-\lambda} (1 \wedge \lambda)^{K_n} \left(\frac{\varepsilon}{K_n}\right)^{cK_n} \end{aligned}$$

for some $c = c(R)$. □

Proof of Lemma 5.2. If $N \sim \text{Pois}(\lambda)$ and $M \geq \max(2\lambda e, 1)$, then, using Stirling's formula,

$$\mathbb{P}(N \geq MK_n) = e^{-\lambda} \sum_{k=\lceil MK_n \rceil}^{\infty} \frac{\lambda^k}{k!} \leq \sum_{k=\lceil MK_n \rceil}^{\infty} \left(\frac{\lambda e}{k}\right)^k \leq \sum_{k=\lceil MK_n \rceil}^{\infty} \left(\frac{1}{2K_n}\right)^k \leq K_n^{-MK_n}.$$

With $t_n := (MK_n L^{-1} \log K_n)^{1/\gamma}$ and the assumption on the tail behavior of the jump heights, we obtain

$$\mathbb{P}\left(\max_{i=0, \dots, N} |\Delta_i| \geq t_n\right) \leq MK_n \mathbb{P}(|\Delta_1| \geq t_n) + \mathbb{P}(N \geq MK_n) \leq (M/L + 1) K_n^{-MK_n}.$$

Define Θ_n as the space of piecewise constant functions f with $|f(0)| \leq t_n$, maximal jump size bounded by t_n and less than MK_n jumps. By the computations above, $\mathbb{P}(X \notin \Theta_n) \leq (M/L + 2) K_n^{-MK_n}$.

Next, we compute the bracketing number of Θ_n with respect to the L^1 -norm. Let $r_{\varepsilon, n}$ be such that $\varepsilon/(4MK_n t_n) \leq r_{\varepsilon, n} \leq \varepsilon/(2MK_n t_n)$ and $1/r_{\varepsilon, n}$ is an integer. Define $x_j := jr_{\varepsilon, n}$ for $0 \leq j < 1/r_{\varepsilon, n}$. In y -direction, consider the grid points $y_\ell := \ell\varepsilon/2$, $\ell = -S_{\varepsilon, n}, \dots, S_{\varepsilon, n}$ with $S_{\varepsilon, n} = \lfloor 2MK_n t_n/\varepsilon \rfloor$. Let $\Theta_n^0 \subset \Theta_n$ be the space of piecewise constant functions in Θ_n with all jumps locations on the grid points x_j , and function values in the discrete set $\{y_\ell : \ell = -S_{\varepsilon, n}, \dots, S_{\varepsilon, n}\}$. We prove that for any function $f \in \Theta_n$, there exists a function $h \in \Theta_n^0$ such that $h \leq f$ and $\|h - f\|_1 \leq \varepsilon$. Consider

$$h = \sum_{j=1}^{1/r_{\varepsilon, n}} \max \left\{ y_\ell : y_\ell \leq \min_{x \in [x_{j-1}, x_j]} f(x) \right\} \mathbf{1}_{[x_{j-1}, x_j]}.$$

Obviously, $h \in \Theta_n^0$ and $h \leq f$. Let us show $\|h - f\|_1 \leq \varepsilon$. Observe that $\|h - \tilde{h}\|_\infty \leq \varepsilon/2$ with $\tilde{h} = \sum_{j=1}^{1/r_{\varepsilon, n}} \min_{x \in [x_{j-1}, x_j]} f(x) \mathbf{1}_{[x_{j-1}, x_j]}$. If f jumps k times on the interval $[x_{j-1}, x_j]$ then $\sup_{x \in [x_{j-1}, x_j]} |f(x) - \tilde{h}(x)| \leq kt_n$. Since the total number of jumps is bounded by MK_n ,

$\|f - \tilde{h}\|_1 \leq MK_n t_n r_{\varepsilon,n} = \varepsilon/2$ implying $\|f - h\|_1 \leq \varepsilon$. There are at most $\binom{r_{\varepsilon,n}}{\ell} (2S_{\varepsilon,n} + 1)^{\ell+1}$ functions in Θ_n^0 with ℓ jumps. The cardinality of Θ_n^0 is therefore bounded by

$$\begin{aligned} \sum_{\ell=0}^{MK_n} \binom{1/r_{\varepsilon,n}}{\ell} (2S_{\varepsilon,n} + 1)^{\ell+1} &\leq \sum_{\ell=0}^{MK_n} r_{\varepsilon,n}^{-\ell} (2S_{\varepsilon,n} + 1)^{\ell+1} \leq 2r_{\varepsilon,n}^{-MK_n} (2S_{\varepsilon,n} + 1)^{MK_n+1} \\ &\leq C'' \left(\frac{K_n}{\varepsilon} \right)^{C'K_n} \end{aligned}$$

for suitable constants C' and C'' . \square

Proof of Proposition 5.4. Let us put $N := \varepsilon^{-1}$ and $f_0(t) := 0$ for $t < 0$ in the sequel. Noting that $X_{i/N} \leq f_0((i-1)/N)$, $i = 1, \dots, N$, implies by monotonicity $X_t \leq f_0(t)$ for all $t \in [0, 1]$, we consider the event $\mathcal{E}_N = \{f_0((i-1)/N) - X_{i/N} \in [0, 2/N], i = 1, \dots, N\}$. On \mathcal{E}_N we deduce from

$$X_t \geq X_{i/N} \geq f_0((i-1)/N) - 2/N \geq f_0(t - 2/N) - 2/N \text{ for } t \in [i/N, (i+1)/N]$$

that $\|X - f_0\|_{L^1} \leq \int_{1-2/N}^1 f_0(s) ds + 2/N \leq (2R+2)/N$ and hence

$$P(\|X - f_0\|_{L^1} \leq (2R+2)/N, X \leq f_0) \geq P(\mathcal{E}_N).$$

Using the independence of increments, we obtain by conditioning on $X_{1/N}$ for intervals I_1, I_2

$$\begin{aligned} P(X_{1/N} \in I_1, X_{2/N} \in I_2) &= \mathbb{E}[1(X_{1/N} \in I_1) P(X_{2/N} - X_{1/N} \in I_2 - x) |_{x=X_{1/N}}] \\ &\geq P(X_{1/N} \in I_1) \inf_{x \in I_1} P(X_{2/N} - X_{1/N} \in I_2 - x). \end{aligned}$$

Using this argument inductively for all $X_{i/N}$ we arrive at

$$\begin{aligned} P(\mathcal{E}_N) &\geq P(f_0(0) - X_{1/N} \in [0, 2/N]) \times \\ &\quad \prod_{j=1}^{N-1} \inf_{h \in [0, 1/N]} P((f_0(j/N) - f_0((j-1)/N)) - (X_{(j+1)/N} - X_{j/N}) \in [h - 2/N, h]). \end{aligned}$$

By assumption, we obtain for $f_0(0) > 2/N$ by considering the case of one jump:

$$\begin{aligned} P(f_0(0) - X_{1/N} \in [0, 2/N]) &\geq P(X \text{ jumps once in } [0, 1/N], X_{1/N} \in [f_0(0) - 1/N, f_0(0)]) \\ &\geq \frac{\lambda}{N} e^{-\lambda/N} \frac{c\rho e^{-\rho f_0(0)}}{N}. \end{aligned}$$

For $f_0(0) \in [0, 2/N]$ we bound by the probability of no jump:

$$P(f_0(0) - X_{1/N} \in [0, 2/N]) \geq P(X_{1/N} = 0) = e^{-\lambda/N}.$$

For $\lambda \leq N$ this lower bound is larger than that for the case $f_0(0) > 2/N$. We conclude for any $f_0(0)$

$$P(f_0(0) - X_{1/N} \in [0, 2/N]) \geq \frac{c\rho\lambda}{N^2} e^{-\lambda/N} e^{-\rho f_0(0)}.$$

For the transition probabilities, consider similarly first $f_0(j/N) - f_0((j-1)/N) \leq 2/N$ and the no jump event, in which case

$$\inf_{h \in [0, 2/N]} P((f_0(j/N) - f_0((j-1)/N)) - (X_{(j+1)/N} - X_{j/N}) \in [h - 2/N, h]) \geq e^{-\lambda/N}.$$

Otherwise $f_0(j/N) - f_0((j-1)/N) > 2/N$ and the probability of the corresponding one jump event implies

$$\begin{aligned} & \inf_{h \in [0, 2/N]} P((f_0(j/N) - f_0((j-1)/N)) - (X_{(j+1)/N} - X_{j/N}) \in [h - 2/N, h]) \\ & \geq \frac{c\rho\lambda}{N^2} e^{-\lambda/N} \exp\left(-\rho(f_0(j/N) - f_0((j-1)/N) + 2/N)\right). \end{aligned}$$

Again, for $\lambda \leq N$ the last bound is also valid for the case $f_0(j/N) - f_0((j-1)/N) \leq 2/N$ and we conclude

$$\begin{aligned} \log(P(\mathcal{E}_N)) & \geq N \log(c\rho\lambda/N^2) - \lambda - \rho \sum_{j=0}^{N-1} (f_0(j/N) - f_0((j-1)/N) + 2/N) \\ & \geq N \log(c\rho\lambda) - 2N \log(N) - \lambda - \rho(R+2). \end{aligned}$$

For $\rho\lambda \gtrsim N^{-p}$ for some $p > 0$ and $\lambda \vee \rho \lesssim N \log(N)$ we infer $\log(P(\mathcal{E}_N)) \gtrsim -N \log(N)$. For $\lambda \asymp \rho \asymp N$, however, we find $\log(P(\mathcal{E}_N)) \gtrsim -N$. \square

Proof of Proposition 5.5. We shall use the following small ball probability of an α -stable subordinator around zero:

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^{\alpha/(1-\alpha)} \log(P(\|X\|_{L^\infty([0,1])} \leq \varepsilon)) \in (-\infty, 0),$$

which follows from the main result in [31] noting that for non-decreasing functions starting in zero the 1-variation equals the supremum norm. This result shows that the α -stable subordinators satisfy the small ball probability in L^∞ with rate $e^{-c\varepsilon^{-1}}$ if and only if $\alpha \leq 1/2$.

Introducing $\nu_<(x) = \nu(x)\mathbf{1}(x < \varepsilon)$, $\nu_>(x) = \nu(x)\mathbf{1}(x \geq \varepsilon)$, we can decompose X as $X^< + X^>$ with two independent Lévy processes $X^<, X^>$ having Lévy densities $\nu_<, \nu_>$, respectively. The small jump process $X^<$ is a subordinator whose Lévy density is smaller than $\nu_{1/2}(x) = C_2 x^{-3/2} \mathbf{1}(x > 0)$, the Lévy density of a stable subordinator $X^{(1/2)}$ of index $\alpha = 1/2$. We can thus couple $X^<$ and $X^{(1/2)}$ such that $X_t^< \leq X_t^{(1/2)}$ holds for all $t \geq 0$ a.s. By the above result, this gives

$$\log(P(\|X^<\|_{L^\infty([0,1])} \leq \varepsilon)) \gtrsim -\varepsilon^{-1}.$$

The process $X^>$, on the other hand, satisfies by the previous proposition (note $f_0(0) \geq \varepsilon$, $\lambda = \int_\varepsilon^\infty \nu \leq \varepsilon^{-1}$):

$$\log \left(P(\|X^> - (f_0 - \varepsilon)\|_{L^1} \leq (2R+2)\varepsilon, X \leq f_0 - \varepsilon) \right) \geq \varepsilon^{-1} \log(c_1 \rho) - 2\varepsilon^{-1} \log(\varepsilon^{-1}) - \varepsilon^{-1} - \rho(R+2).$$

By independence, we conclude for $X = X^< + X^>$:

$$\begin{aligned} & \log \left(P(\|X - f_0\|_{L^1} \leq (2R+3)\varepsilon, X \leq f_0) \right) \\ & \geq \log \left(P(\|X^> - (f_0 - \varepsilon)\|_{L^1} \leq (2R+1)\varepsilon, X^> \leq f_0 - \varepsilon, X^< \leq \varepsilon) \right) \\ & \gtrsim -\varepsilon^{-1} \log(\varepsilon^{-1}) - \varepsilon^{-1}. \end{aligned}$$

This gives the result. \square

Proof of Proposition 5.6. For $n \geq 16$ there exists a positive integer N with $\frac{1}{4}\sqrt{n} \leq N \leq \frac{1}{2}\sqrt{n}$. Define $I_j := [(j-1)/N, j/N)$ and consider the piecewise constant lower approximation ℓ_f of $(f - f_0)_+$ with values in $N^{-1}\mathbb{Z}$:

$$\ell_f(x) := \min_{x \in I_j} \lfloor N(f - f_0)_+(x) \rfloor / N \text{ for } x \in I_j.$$

Then $\ell_f(x) \geq (f \vee f_0)(x - 1/N) - f_0(x + 1/N) - 1/N$ by monotonicity of $f, f_0, f \vee f_0$ (with constant extensions of $(f \vee f_0)(x) = (f \vee f_0)(0)$ for $x < 0$, $f_0(x) = f_0(1)$ for $x > 1$). Let $\varepsilon_n = (4R+8)/\sqrt{n}$. From $\int_0^{1-1/N} (f \vee f_0 - f_0) \geq \int_0^{1-4/\sqrt{n}} (f \vee f_0 - f_0) \geq \varepsilon_n$ we deduce

$$\int_0^1 \ell_f \geq \int_0^{1-1/N} (f \vee f_0) + \frac{(f \vee f_0)(0)}{N} - \int_{1/N}^1 f_0 - \frac{f_0(1)}{N} - \frac{1}{N} \geq \varepsilon_n - \frac{1 + f_0(1) - f_0(0)}{N}.$$

Since $N \geq (R+2)\varepsilon_n^{-1}$ we arrive at $\int \ell_f \geq N^{-1}$.

All these functions ℓ_f are therefore lower bounded by the functions $\ell^a(x) = N^{-1} \sum_{j=1}^N a_j \mathbf{1}_{I_j}$ with $a \in \mathbb{N}_0^N$, $\sum_j a_j = N$. A simple combinatorial argument shows

$$|\{a \in \mathbb{N}_0^N \mid \sum_j a_j = N\}| = \frac{N^N}{N!} \leq e^N.$$

Using the family $(f_0 + \ell^a)_a$ of lower bound functions for $f \vee f_0$, we therefore have

$$S\left(n, \left\{f \in \Theta_n \mid \int_0^{1-4/\sqrt{n}} (f - f_0)_+ \geq \varepsilon_n\right\}, f_0\right) \leq \sum_a e^{-n \int \ell^a} \leq e^{N-nN^{-1}} \leq e^{-\sqrt{n}}.$$

\square

10.1 Proofs for Section 6

Proof of Theorem 6.1. Set $r_{j,n} := n(t_j - t_{j-1})$, $r_n = \min_j r_{j,n}$, $c := \inf_{-2R \leq x \leq 2R} g(x)$. Since g is continuous and positive, c is positive. Denote by $\|g\|_{\mathcal{C}^\beta}$ the Hölder β -norm. The total variation distance between a probability measure P and the conditional probability measure $P(\cdot|A)$ is bounded by $2P(A^c)$. We apply this to the posterior distribution and Q^n for the set

$$A = \left\{ 0 \leq \max_j r_{j,n} (\hat{a}_j - a_j)_+ \leq 2 \log K_n \right\}.$$

It then remains to prove that uniformly over $f_0 \in \text{PC}^*(K_n, (t_0, \dots, t_{K_n}), R)$,

$$(i) \ E_{f_0}[\|\Pi(\cdot \cap A|N)/\Pi(A|N) - Q^n(\cdot|A)\|_{\text{TV}}] \rightarrow 0$$

$$(ii) \ E_{f_0}[\Pi(A^c|N) + Q^n(A^c)] \rightarrow 0.$$

Proof of (i): Recall that $-R \leq a_j^0 \leq \hat{a}_j$. Then on the event $\{\max_j \hat{a}_j \leq 3R/2\} \cap A$

$$g(a_j) \leq g(\hat{a}_j) + \|g\|_{\mathcal{C}^\beta} \left(\frac{2 \log K_n}{r_n} \right)^\beta \leq g(\hat{a}_j) (1 + R_n) \text{ with } R_n = \frac{\|g\|_{\mathcal{C}^\beta}}{c} \left(\frac{2 \log K_n}{r_n} \right)^\beta.$$

Similarly, we find $g(a_j) \geq g(\hat{a}_j) (1 - R_n)$. For an arbitrary Borel set B and $f_{\mathbf{a}} = \sum_{j=1}^K a_j \mathbf{1}_{[t_{j-1}, t_j]}$ we obtain

$$\begin{aligned} \frac{\Pi(B \cap A|N)}{\Pi(A|N)} &= \frac{\int_{B \cap A} e^{n \int f_{\mathbf{a}}} \mathbf{1}(\forall i : f_{\mathbf{a}}(X_i) \leq Y_i) \prod_j g(a_j) d\mathbf{a}}{\int_A e^{n \int f_{\mathbf{a}}} \mathbf{1}(\forall i : f_{\mathbf{a}}(X_i) \leq Y_i) \prod_j g(a_j) d\mathbf{a}} \\ &\leq \frac{\sup_{(a_1, \dots, a_{K_n}) \in A} \prod_j g(a_j)}{\inf_{(a_1, \dots, a_{K_n}) \in A} \prod_j g(a_j)} Q^n(B|A) \\ &\leq \left(\frac{1 + R_n}{1 - R_n} \right)^{K_n} Q^n(B|A). \end{aligned}$$

By assumption, there is some N_0 such that $R_n \leq 1/2$ for all $n \geq N_0$. This gives

$$\left(\frac{1 + R_n}{1 - R_n} \right)^{K_n} \leq (1 + 4R_n)^{K_n} \leq \exp(4R_n K_n),$$

which proves that

$$\sup_B \left(\frac{\Pi(B \cap A|N)}{\Pi(A|N)} - Q^n(B|A) \right) \leq \exp(4R_n K_n) - 1. \quad (10.1)$$

Similar arguments together with the Bernoulli inequality imply

$$\frac{\Pi(B \cap A|N)}{\Pi(A|N)} \geq \left(\frac{1 - R_n}{1 + R_n} \right)^{K_n} Q^n(B|A) \geq (1 - R_n)^{K_n} Q^n(B|A) \geq Q^n(B|A) - R_n K_n,$$

whenever $n \geq N_0$. Together with (10.1) and the assumption $R_n K_n \rightarrow 0$ this gives

$$\|\Pi(\cdot \cap A|N)/\Pi(A|N) - Q^n(\cdot|A)\|_{\text{TV}} \mathbf{1}(\max_j \hat{a}_j \leq 3R/2) \rightarrow 0. \quad (10.2)$$

Notice that under P_{f_0} , $\hat{a}_j - a_0 \sim \text{Exp}(r_{j,n})$. Thus with $\xi_j \sim \text{Exp}(r_{j,n})$,

$$P_{f_0}\left(\max_j \hat{a}_j \geq 3R/2\right) \leq \sum_{j=1}^{K_n} \mathbb{P}(\xi_j \geq R/2) \leq K_n e^{-r_n R/2} \rightarrow 0. \quad (10.3)$$

Together with (10.2), this proves (i).

Proof of (ii) : The density of Q^n factorizes as $\prod_{j=1}^{K_n} r_{j,n} e^{r_{j,n}(a_j - \hat{a}_j)} \mathbf{1}(a_j \leq \hat{a}_j)$. By a union bound we obtain

$$Q^n\left(\max_j r_{j,n}(\hat{a}_j - a_j)_+ \geq 2 \log K_n\right) \leq \sum_{j=1}^{K_n} Q^n((\hat{a}_j - a_j)_+ \geq 2r_{j,n}^{-1} \log K_n) = \frac{1}{K_n} \rightarrow 0.$$

In the second part of the proof we shall show that for sufficiently large n ,

$$\max_j E_{f_0}[\Pi((\hat{a}_j - a_j)_+ \geq 2r_{j,n}^{-1} \log K_n | N)] \lesssim \frac{1}{K_n^2}.$$

which together with the union bound completes the proof for (ii).

Since the likelihood factorizes as $e^{n \int f_{\mathbf{a}} \mathbf{1}(\forall i : f_{\mathbf{a}}(X_i) \leq Y_i)} = \prod_{j=1}^{K_n} e^{r_{j,n} a_j} \mathbf{1}(a_j \leq \hat{a}_j)$, we find, using $\|g\|_{\infty} \leq \|g\|_{\mathcal{C}^{\beta}}$,

$$\begin{aligned} \Pi((\hat{a}_j - a_j)_+ \geq 2r_{j,n}^{-1} \log K_n | N) &= \frac{\int_{-\infty}^{\hat{a}_j - 2 \log(K_n)/r_{j,n}} e^{r_{j,n} a_j} g(a_j) da_j}{\int_{-\infty}^{\hat{a}_j} e^{r_{j,n} a_j} g(a_j) da_j} \\ &\leq \frac{\|g\|_{\mathcal{C}^{\beta}} e^{r_{j,n} \hat{a}_j}}{K_n^2 r_{j,n} \int_{-\infty}^{\hat{a}_j} e^{r_{j,n} a_j} g(a_j) da_j}. \end{aligned}$$

Recall that $|a_j^0| \leq R$ and $a_j^0 \leq \hat{a}_j$. As in (i) we work on the event $\hat{a}_j \leq 3R/2$. Then in the denominator we can bound from below

$$\int_{-\infty}^{\hat{a}_j} e^{r_{j,n} a_j} g(a_j) da_j \geq c \int_{\hat{a}_j - R/2}^{\hat{a}_j} e^{r_{j,n} a_j} da_j \geq \frac{c}{r_{j,n}} e^{r_{j,n} \hat{a}_j} (1 - e^{-r_{j,n} R/2}).$$

Let N'_0 such that $r_n \geq 2/R$ for all $n \geq N'_0$. Then, for $n \geq N'_0$,

$$\Pi((\hat{a}_j - a_j)_+ \geq 2r_{j,n}^{-1} \log K_n | N) \mathbf{1}(\hat{a}_j \leq 3R/2) \leq \frac{\|g\|_{\mathcal{C}^{\beta}}}{K_n^2 c (1 - e^{-1})}.$$

Together with (10.3) and $K_n e^{-r_n R/2} \lesssim 1/K_n^2$, this yields

$$\max_j E_{f_0}[\Pi((\hat{a}_j - a_j)_+ \geq 2r_{j,n}^{-1} \log K_n | N)] \lesssim \frac{1}{K_n^2}.$$

This shows (ii) and completes the proof. \square

Proof of Corollary 6.2. By a slight abuse of notation we write $\|P_X - P\|_{\text{TV}}$ as $\text{TV}(X, P)$ when $X \sim P_X$. By Theorem 6.1, the marginal posterior of ϑ converges under P_{f_0} in total variation to the distribution $\int \hat{f}^{\text{MLE}} - \sum_{j=1}^{K_n} (t_j - t_{j-1}) \eta_j$ with $\eta_j \sim \text{Exp}(n(t_j - t_{j-1}))$. Thus, $(t_j - t_{j-1}) \eta_j \sim n^{-1} \text{Exp}(1)$. Invertible transformations do not change the total variation distance. For independent $\xi_j \sim \text{Exp}(1)$ we therefore get

$$\text{TV} \left(\int \hat{f}^{\text{MLE}} - \frac{\sum_j \xi_j}{n}, \mathcal{N} \left(\int \hat{f}^{\text{MLE}} - \frac{K_n}{n}, \frac{K_n}{n^2} \right) \right) = \text{TV} \left(K_n^{-1/2} \sum_{j=1}^{K_n} (\xi_j - 1), \mathcal{N}(0, 1) \right).$$

By the CLT in total variation (cf. [1], Theorem 2.5), the latter converges to zero as $K_n \rightarrow \infty$. This completes the proof of the first assertion. It also implies that $I(\alpha)$ is an asymptotic $(1 - \alpha)$ -credible interval.

It remains to prove that $I(\alpha)$ is also an honest confidence interval. By the explicit law of \hat{f}^{MLE} , we conclude that under P_{f_0} , $\int \hat{f}^{\text{MLE}} = \int f_0 + n^{-1} \sum_{j=1}^{K_n} \xi'_j$ with independent $\xi'_j \sim \text{Exp}(1)$. Thus, uniformly in f_0 it holds

$$P_{f_0} \left(\int f_0 \in I(\alpha) \right) = \mathbb{P} \left(\Phi^{-1}(\alpha/2) \leq K_n^{-1/2} \sum_{j=1}^{K_n} (\xi'_j - 1) \leq \Phi^{-1}(1 - \alpha/2) \right) \rightarrow 1 - \alpha,$$

using the standard central limit theorem. \square

Proof of Proposition 6.3. The first assertion follows from a simple PPP probability calculation. Let us derive bounds for the expectation and the second moment of V_{jn} . Let $r > 0$. The formula $\int_0^1 y e^{-ry^2} dy = (1 - e^{-r})/(2r)$ and integration by parts gives the identity

$$\int_0^1 e^{-ry^2} dy + \frac{1}{2r} e^{-r} = \frac{1}{2r} + \int_0^1 \int_0^z e^{-ry^2} dy dz.$$

With $r = n/(2K_n^2)$, $E[V_{jn}]$ can therefore be rewritten as

$$\begin{aligned} E[V_{jn}] &= \int_0^\infty P(V_{jn} \geq y) dy = \int_0^1 e^{-\frac{n}{2K_n^2} y^2} dy + \frac{K_n^2}{n} e^{-\frac{n}{2K_n^2}} = \int_0^1 \int_0^z e^{-\frac{n}{2K_n^2} y^2} dy dz + \frac{K_n^2}{n} \\ &\leq \frac{3}{8} + \frac{1}{8} e^{-\frac{n}{8K_n^2}} + \frac{K_n^2}{n} \leq \frac{7}{16} \vee \left(\frac{1}{2} - \frac{n}{2^7 K_n^2} \right) + \frac{K_n^2}{n}, \end{aligned} \quad (10.4)$$

where for the first inequality, we decomposed the double integral into $\int_0^1 = \int_0^{1/2} + \int_{1/2}^1$ and $\int_0^z = \int_0^{1/2} + \int_{1/2}^z$ for $z \geq 1/2$ and for the second inequality used $e^{-x} \leq 1/2 \vee (1 - x/2)$ for $x \geq 0$. Moreover,

$$E[V_{jn}^2] = \int_0^\infty P(V_{jn} \geq \sqrt{y}) dy = \int_0^1 e^{-\frac{n}{2K_n^2} y} dy + e^{\frac{n}{2K_n^2}} \int_1^\infty e^{-\frac{n}{K_n^2} \sqrt{y}} dy$$

$$= 2\frac{K_n^2}{n}(1 - e^{-\frac{n}{2K_n^2}}) + 2e^{\frac{n}{2K_n^2}} \int_1^\infty v e^{-\frac{n}{K_n^2}v} dv \leq 2\frac{K_n^2}{n} + 8\frac{K_n^4}{n^2}. \quad (10.5)$$

Because of (10.4) and $\hat{a}_j = a_j^0 + (j - 1 + V_{jn})/K_n$,

$$\begin{aligned} E_{f_0} \left[\int \hat{f}^{\text{MLE}} - \frac{K_n}{n} \right] &= \frac{1}{K_n} \sum_{j=1}^{K_n} E_{f_0}[\hat{a}_j] - \frac{K_n}{n} = \int f_0 - \frac{1}{2} + \frac{K_n(K_n - 1)}{2K_n^2} + \frac{E[V_{1n}]}{K_n} - \frac{K_n}{n} \\ &\leq \int f_0 - \frac{n}{2^7 K_n^3} \end{aligned}$$

and this completes the proof. \square

Proof of Lemma 6.4. A close inspection of the proof shows that Theorem 2.1 in [28] also holds for functions which are $C^\beta(R)$ on each interval $[kh, k(h + 1))$. Define $I_k := ((k - 1)/K_n, k/K_n]$, $Y_k^* := \min_{i: X_i \in I_k} Y_i$,

$$\hat{\vartheta}_k := \left(Y_k^* + \frac{1}{K_n} \right) - \frac{K_n}{n} \sum_{i \geq 1} \mathbf{1} \left(X_i \in I_k, Y_i \leq Y_k^* + \frac{1}{K_n} \right),$$

and $\hat{\vartheta}^{\text{block}} = K_n^{-1} \sum_{k=1}^{K_n} \hat{\vartheta}_k$. To obtain the expectation and a bound on the variance of $\hat{\vartheta}^{\text{block}}$, we can apply Theorem 2.1 in [28] with $w = 1, \beta = R = 1$ and $h = 1/K_n$ since the true support boundary function is in Lip_{K_n} . This gives $E_{\vartheta_0}[\hat{\vartheta}^{\text{block}}] = \vartheta_0$ and $\text{Var}(\hat{\vartheta}^{\text{block}}) = 2/(K_n n) + K_n/n^2$. If $C(\alpha) = [\vartheta - \alpha^{-1/2}(2/(K_n n) + K_n/n^2)^{1/2}, \vartheta + \alpha^{-1/2}(2/(K_n n) + K_n/n^2)^{1/2}]$, then

$$P_{f_0}(\vartheta_0 \notin C(\alpha)) \geq P_{f_0}(|\hat{\vartheta}^{\text{block}} - \vartheta_0| \leq \alpha^{-1/2} \text{Var}_{f_0}(\hat{\vartheta}^{\text{block}})^{1/2}) \leq \alpha$$

follows from Chebyshev's inequality. The length of $C(\alpha)$ is $O(\sqrt{K_n}/n + 1/\sqrt{K_n n})$. \square

Proof of Theorem 6.5. We first prove the Bernstein-von Mises type result $E_{f_0}[|\Pi(\cdot|N) - Q^n|_{\text{TV}}] \rightarrow 0$. For $(a_1, \dots, a_{K_n}) \sim Q^n$ we have

$$Q^n(\min_j a_j < -R) \leq K_n \int_{-\infty}^{-R} \frac{K_n}{n} e^{\frac{n}{K_n}(a_j - \hat{a}_j)} da_j \leq K_n e^{-\frac{n}{K_n}R} \leq K_n n^{-R} \rightarrow 0.$$

Arguing as in the proof of Theorem 6.1, it follows that

$$\|Q^n - Q^n(\cdot | \min_j a_j \geq -R)\|_{\text{TV}} \leq 2Q^n(\min_j a_j < -R) \rightarrow 0. \quad (10.6)$$

On the event $\mathcal{A} := \{\max_j \hat{a}_j \leq R\}$ we have equality

$$\Pi((a_1, \dots, a_K) \in B|N) = \frac{\int_{B \cap [-R, R]^{K_n}} e^{\frac{n}{K_n} \sum_j a_j} \mathbf{1}(\forall j : a_j \leq \hat{a}_j) d\mathbf{a}}{\int_{[-R, R]^{K_n}} e^{\frac{n}{K_n} \sum_j a_j} \mathbf{1}(\forall j : a_j \leq \hat{a}_j) d\mathbf{a}} = Q^n(B | \min_j a_j \geq -R).$$

Thus,

$$E_{f_0}[\|\Pi(\cdot|N) - Q^n(\cdot|\min_j a_j \geq -R)\|_{\text{TV}}] \leq P_{f_0}(\mathcal{A}^c). \quad (10.7)$$

By Proposition 6.3, $\hat{a}_j = (j - 1 + V_{jn})/K_n$, in distribution and thus

$$P_{f_0}(\mathcal{A}^c) \leq K_n P_{f_0}(V_{jn} \geq K_n(R - 1)) \leq K_n e^{-\frac{n}{2K_n}(R-1)} \leq K_n n^{-(R-1)/2} \rightarrow 0,$$

where the last step follows because of $R > 3$. With (10.6) and (10.7), we obtain the Bernstein-von Mises type result $E_{f_0}[\|\Pi(\cdot|N) - Q^n\|_{\text{TV}}] \rightarrow 0$. Arguing as in the proof of Corollary 6.2, we can then conclude that for the marginal posterior of $\vartheta = \int f$,

$$E_{f_0} \left[\left\| \Pi(\vartheta \in \cdot | N) - \mathcal{N} \left(\int \hat{f}^{\text{MLE}} - \frac{K_n}{n}, \frac{K_n}{n^2} \right) \right\|_{\text{TV}} \right] \rightarrow 0.$$

This proves $E_{f_0}[\Pi(\vartheta \in I(\alpha) | N)] \rightarrow 1 - \alpha$.

We turn to proving (6.5). Recall that $f_0(x) = x$ and $a_n = 2^{-8}(nK_n^{-3/2} \wedge n^2K_n^{-7/2})$. With $\sigma_n^2 := \text{Var}_{f_0}(V_{1n})$ and

$$A_n := \frac{K_n^{3/2}}{\sigma_n} \left(\frac{1}{2K_n} + \frac{K_n}{n} - \frac{E[V_{1n}]}{K_n} - \frac{\sqrt{K_n}}{n} a_n \right),$$

we obtain by Chebyshev inequality

$$P_{f_0} \left(\frac{1}{2} \leq \int \hat{f}^{\text{MLE}} - \frac{K_n}{n} + \frac{\sqrt{K_n}}{n} a_n \right) = P_{f_0} \left(\frac{\sum_{j=1}^{K_n} V_{j,n} - E[V_{j,n}]}{\sqrt{K_n} \sigma_n} \geq A_n \right) \leq \frac{1}{A_n^2}, \quad (10.8)$$

If $K_n \leq \sqrt{n/8}$, then, with (10.4) and (10.5),

$$A_n \geq \frac{K_n^{3/2}}{\sigma_n} \left(\frac{1}{16K_n} - \frac{\sqrt{K_n}}{n} a_n \right) \rightarrow \infty.$$

On the other hand, if $\sqrt{n/8} \leq K_n = o(n^{4/7})$, with (10.4) and (10.5),

$$A_n \geq \frac{K_n^{3/2}}{\sigma_n} \left(\frac{n}{2^7 K_n^3} - \frac{\sqrt{K_n}}{n} a_n \right) \rightarrow \infty.$$

Together with (10.8) this proves (6.5). The last claim follows from the definition of $I(\alpha)$ in (6.3) and the fact that $a_n \rightarrow \infty$ for $K_n = o(n^{4/7})$. \square

References

- [1] BALLY, V., AND CARAMELLINO, L. Asymptotic development for the CLT in total variation distance. *Bernoulli* 22, 4 (2016), 2442–2485.

- [2] BOCHKINA, N. A., AND GREEN, P. J. The Bernstein - von Mises theorem and nonregular models. *Ann. Statist.* 42, 5 (2014), 1850–1878.
- [3] CASTILLO, I., AND NICKL, R. Nonparametric Bernstein-von Mises theorems in Gaussian white noise. *Ann. Statist.* 41, 4 (2013), 1999–2028.
- [4] CASTILLO, I., AND NICKL, R. On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* 42, 5 (2014), 1941–1969.
- [5] CASTILLO, I., SCHMIDT-HIEBER, J., AND VAN DER VAART, A. Bayesian linear regression with sparse priors. *Ann. Statist.* 43, 5 (2015), 1986–2018.
- [6] CASTILLO, I., AND VAN DER VAART, A. W. Needles and straw in a haystack: posterior concentration for possibly sparse sequences. *Ann. Statist.* 40, 4 (2012), 2069–2101.
- [7] CHERNOZHUKOV, V., AND HONG, H. Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* 72, 5 (2004), 1445–1480.
- [8] COHEN, A., DAUBECHIES, I., AND VIAL, P. Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* 1, 1 (1993), 54–81.
- [9] GHOSAL, S., GHOSH, J. K., AND VAN DER VAART, A. W. Convergence rates of posterior distributions. *Ann. Statist.* 28, 2 (2000), 500–531.
- [10] GHOSAL, S., AND VAN DER VAART, A. Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* 35, 1 (2007), 192–223.
- [11] GINÉ, E., AND NICKL, R. Rates on contraction for posterior distributions in L^r -metrics, $1 \leq r \leq \infty$. *Ann. Statist.* 39, 6 (2011), 2883–2911.
- [12] GINÉ, E., AND NICKL, R. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, Cambridge, 2016.
- [13] HERMANSEN, G. H., AND HJORT, N. L. Bernshtein–von Mises theorems for nonparametric function analysis via locally constant modelling: a unified approach. *J. Statist. Plann. Inference* 166 (2015), 138–157.
- [14] HOFFMANN, M., ROUSSEAU, J., AND SCHMIDT-HIEBER, J. On adaptive posterior concentration rates. *Ann. Statist.* 43, 5 (2015), 2259–2295.
- [15] JIRAK, M., MEISTER, A., AND REISS, M. Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* 42, 5 (2014), 1970–2002.
- [16] KIM, Y., AND LEE, J. A Bernstein - von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.* 32, 4 (2004), 1492–1512.

- [17] KLEIJN, B., AND KNAPIK, B. Semiparametric posterior limits under local asymptotic exponentiality. *ArXiv e-prints* (2012).
- [18] KLEIJN, B. J. K., AND VAN DER VAART, A. W. The Bernstein-Von-Mises theorem under misspecification. *Electron. J. Stat.* 6 (2012), 354–381.
- [19] KOROSTEL'EV, A. P., AND TSYBAKOV, A. B. *Minimax theory of image reconstruction*, vol. 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1993.
- [20] KUTOYANTS, Y. *Statistical Inference for Spatial Poisson Processes*. Springer, 1998.
- [21] LEAHU, H. On the Bernstein-von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.* 5 (2011), 373–404.
- [22] LI, M., AND GHOSAL, S. Bayesian detection of image boundaries. *Ann. Statist.* (2017). to appear.
- [23] LO, A. Y. Bayesian nonparametric statistical inference for Poisson point processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 59, 1 (1982), 55–66.
- [24] MEISTER, A., AND REISS, M. Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Theory Related Fields* 155, 1-2 (2013), 201–229.
- [25] PANOV, M., AND SPOKOINY, V. Finite sample Bernstein–von Mises theorem for semiparametric problems. *Bayesian Anal.* 10, 3 (2015), 665–710.
- [26] RAY, K. Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* 7 (2013), 2516–2549.
- [27] RAY, K. Adaptive Bernstein-von Mises theorems in Gaussian white noise. *Ann. Statist.* (2017). to appear.
- [28] REISS, M., AND SELK, L. Efficient estimation of functionals in nonparametric boundary models. *Bernoulli* 23, 2 (2017), 1022–1055.
- [29] SALOMOND, J.-B. Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electron. J. Statist.* 8, 1 (2014), 1380–1404.
- [30] SATO, K.-I. *Lévy processes and infinitely divisible distributions*, vol. 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013.
- [31] SIMON, T. Small ball estimates in p -variation for stable processes. *J. Theoret. Probab.* 17, 4 (2004), 979–1002.

- [32] SZABÓ, B., VAN DER VAART, A. W., AND VAN ZANTEN, J. H. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* *43*, 4 (2015), 1391–1428.
- [33] VAN DER VAART, A. W. *Asymptotic statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [34] VAN DER VAART, A. W., AND VAN ZANTEN, H. Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* *36*, 3 (2008), 1435–1463.
- [35] WANG, Y. Small ball problem via wavelets for Gaussian processes. *Statist. Probab. Lett.* *32*, 2 (1997), 133 – 139.